Hemerocallis Species, Hybrids, and Genetics

Terrence P. McGarty

Copyright © 2009 Terrence P. McGarty, all rights reserved.

Contents

1 Introduction 13
1.1 Hybridizing 13
1.2 A Platform for Discovery 17
1.3 Outline of Book 19
1.3.1 Form and Species 19
1.3.2 Genes and Gene Expression 21
1.3.3 Phylogenetics and Classification 21
1.3.4 Flower Color and Color Expression 22
1.3.5 Flower Color and Patterning 23
1.3.6 Classic Genetics and Hybridizing 23
1.3.7 Hybridizers 24
2 Form and Species 25
2.1 Genotype and Phenotype 25
2.2 Species 28
2.3 Breeding and Hybridizing 31
2.4 Classification 32
2.5 Form 33
2.5.1 Parts of the Plant 34
2.5.2 Roots 36
2.5.3 Scapes 37
2.5.4 Branching 37
2.5.5 Buds 41
2.5.6 Flowers 42
2.5.7 Stamens and Pollen 43
2.5.8 Pods 44
2.5.9 Seeds 47
2.6 Species 49
2.6.1 Buds 63
2.6.2 Pods 65
2.6.3 Scapes 67
2.6.4 Seeds 68
2.7 Early Bloomers 70
2.8 Later Blooms 73
2.9 Keys and Classification 75
2.9.1 Simple Key to Twelve Species 75

2.9.2 Stout Key ... 77
2.9.3 Keys to New Species ... 79
2.9.4 Erhardt Key ... 81
2.9.5 Plodeck Key ... 82
2.9.6 Matsuoka and Hotta (1966) Lists Species ... 82
2.10 Appendix: Summary of Species ... 84
2.10.1 Flowers ... 85
2.10.2 Buds ... 87
2.10.3 Pods ... 89
2.10.4 Leaves ... 91
2.10.5 Roots ... 94
2.10.6 Branching ... 96
2.10.7 Seeds ... 99
3 Genetic Principles and Applications ... 106
3.1 Introduction ... 106
3.2 Preliminary Concepts and Definitions ... 106
3.2.1 Chromosomes and Genes ... 107
3.2.2 Chromosome ... 108
3.2.3 DNA ... 110
3.2.4 Gene ... 115
3.3 Genotype and Phenotype ... 117
3.4 Genetics ... 124
3.4.1 Plant Cells ... 124
3.4.2 Plant DNA ... 127
3.4.3 Plant Gene Processes ... 128
3.4.4 Activators ... 131
3.4.5 Repressors ... 134
3.4.6 Summary of Actions ... 135
3.5 Expression Analysis and Implications ... 136
3.5.1 Approach: Engineering versus Science ... 136
3.5.2 A Control Paradigm ... 136
3.5.3 Cell Signaling: Intra and Inter Cell ... 137
3.6 Flower Color Expression ... 139
3.6.1 Pathways and Enzymes ... 140
3.6.2 Anthocyanins ... 141
3.6.3 Other Color Elements ... 144
3.6.4 Carotenoids ... 145
3.6.5 Flavones ... 146
3.6.6 Pathways ... 147
3.6.6.1 Anthocyanin Pathway ... 148
3.6.6.2 Carotenoid Pathway ... 153

3.6.6.3 Flavonol Pathway 154
3.7 Conclusions 154
4 Phylogenetics, DNA, Classification 155
4.1 Introduction 155
4.2 Key Questions 155
4.3 Prior Efforts 157
4.4 The Problem of Classification 160
4.4.1 Morphological Classification 162
4.4.2 Genetic Classification 165
4.5 Genetic Techniques 165
4.5.1 Genes and Restriction Enzymes 165
4.5.1.1 Restriction Enzymes 165
4.5.1.2 PCR 168
4.5.2 Procedures 171
4.5.2.1 RFLP 171
4.5.2.2 Microsatellite 172
4.5.2.3 RAPD 173
4.5.2.4 AFLP 173
4.5.2.5 Microarrays 176
4.5.3 Comparisons 180
4.6 Classification Techniques 181
4.6.1 Principles 181
4.6.2 Measurements and Metrics 182
4.6.3 Techniques for Trees 185
4.6.4 Neighbor Joining 186
4.6.5 Maximum Likelihood 189
4.7 Application to Hemerocallis 198
4.8 Genetic Analysis of Species 203
4.8.1 The Genetic Model 204
4.8.2 Quasi-Speciation 205
4.8.3 Genetic Change 212
4.9 Conclusions 215
5 Flower Color and Anthocyanins 218
5.1 Introduction 218
5.2 Classic Color Theory 219
5.3 Spectra and Measurements 226
5.3.1 Classic Spectrometer 226
5.3.2 Fourier Transform Spectrometer 227

5.4 Plants, Color and Chemistry 229
5.4.1 Molecular Issues 230
5.4.2 Absorbance and Reflectance 231
5.5 Plant Cell Reflection 233
5.6 Colorants from Secondary Pathways 237
5.6.1 Anthocyanins 237
5.6.2 Other Color Elements 241
5.6.2.1 Carotenoids 242
5.6.2.2 Flavones 243
5.6.3 Anthocyanin Absorbance and Reflectance 244
5.7 Estimating Anthocyanin Concentrations 246
5.7.1 The Model 247
5.7.2 The Approaches 249
5.7.2.1 Newton Steepest Descent 250
5.7.2.2 Kalman Filter 252
5.7.2.3 The Matched Filter Approach 255
5.8 Conclusions 257
6 Flower Color, Patterning and Control 259
6.1 Patterning 259
6.1.1 Prior Work on Patterning 260
6.1.2 The Turing Model 261
6.1.3 Experimental Methods for Validation 265
6.1.4 Experimental Results 266
6.1.5 Discussion of Experimental Results 268
6.2 Gene Expression in Plants: Use of System Identification for Control of Color 286
6.2.1 System Models for Gene Expression: 287
6.2.2 Prior Work 288
6.2.2.1 Genetic Structure of Hemerocalis 288
6.2.2.2 Gene Expression and Pathway Control 289
6.2.2.3 Modeling of Gene Expression: Analysis and Synthesis 289
6.2.3 Flower Color Expression 290
6.2.3.1 Pathways and Enzymes 290
6.2.3.2 Anthocyanins 291
6.2.4 Other Color Elements 293
6.2.5 Pathways 294
6.2.5.1 Anthocyanin Pathway 295
6.2.5.2 Carotenoid Pathway 297
6.2.5.3 Flavonol Pathway 297
6.2.6 Expression Analysis and Implications 298
6.2.6.1 Approach: Engineering versus Science 298

6.2.6.2 A Control Paradigm ... 298
6.2.6.3 A Model for Secondary Production ... 299
6.2.7 System Identification ... 306
6.2.7.1 Modification for on/off A/R Genes ... 310
6.2.8 Estimation versus Identification ... 312
6.2.8.1 The Model ... 312
6.2.8.2 The Estimator Model ... 314
6.2.8.3 Model Variants ... 316
6.2.9 Conclusion ... 317
7 Classic Genetics and Hybridizing of Hemerocallis ... 319
7.1 Introduction ... 319
7.2 Classic Mendellian Concepts ... 320
7.3 Basic Mendellian Genetic Analysis ... 324
7.3.1 Simple Crosses ... 326
7.3.2 Complex Crosses ... 330
7.4 Genes, Dominance, Color ... 331
7.4.1 Case 1 YY Parent Self Crossed ... 336
7.4.2 Case 2 yY Parent Self Crossed ... 337
7.5 Summary of Mendellian Approach ... 340
7.5.1 Example ... 341
7.5.1.1 Case 1: Hyperion and Species ... 341
7.5.1.2 Case 2: Bicolors ... 344
7.6 Implications of Mendellian Crosses ... 345
7.6.1 Heritability ... 345
7.6.2 Creating a Homozygous Line ... 346
7.6.3 Heterozygosity and Dominance ... 347
7.6.4 Convergence of Homozygosity ... 349
7.7 Hybridizing or Breeding Techniques ... 352
7.8 Methods and Goals of Crossing ... 353
7.9 Selection Methods ... 353
7.10 Characterizing Goals ... 354
7.11 Methods Applied to Crossing ... 355
7.12 Crossing Methods ... 359
7.12.1 Backcross ... 359
7.12.2 Testcross ... 360
7.12.3 Outcrossing ... 360
7.12.4 Line Breeding ... 363
7.12.5 Mass Selection ... 364
7.12.6 Recurrent Selection ... 365

7.13 Backcrossing: Analysis and Statistical Validation 368
7.14 Statistical Analyses 378
7.15 Discussion 381
7.16 Comparison of Methods 382
7.17 Hybridizing Examples 384
7.18 Bi-Color and Spider 384
7.18.1 F2 Bicolor 385
7.18.2 F2-F3 Eyezones 386
7.18.3 F2 Eyezone and Color Change 387
7.19 Bicolor and Dominance 388
7.20 Blending versus Dominance 389
7.21 Conclusions 391
8 History of Hybridizing 393
8.1 Early Hybridizing Developments 393
8.1.1 Stout 393
8.1.2 Others 395
8.2 Middle Ages of Hybridizing 398
8.2.1 Childs 398
8.2.2 Hall 399
8.2.3 Marsh 400
8.2.4 Peck 401
8.2.5 Winniford 402
8.3 Wild 403
8.4 Recent Hybridizers 404
8.4.1 Stamile, Patrick and Grace 407
8.4.2 Kirchhoff 409
8.4.3 Moldovan 410
8.4.4 Matzek 413
8.4.5 Apps 414
8.4.6 Stevens 419
8.4.7 Davidson 421
8.4.8 Petit 422
8.4.9 Hanson 426
8.4.10 Mahieu 427
8.4.11 Joiner 428
8.4.12 McGarty 428

9 Conclusions 431
9.1 Key Observations.. 431
9.1.1 Genus.. 431
9.1.2 Methods and Procedures and Processes 432
9.2 Unanswered Issues 432
9.2.1 Genus.. 432
9.2.2 Methods and Procedures and Processes 433
9.3 Extensions 433
10 References 434

To my wife Sara….

for her many days of help in the fields of lilies….

and for my first collection of lilies. Without the love and help none of this would have been achievable.

Preface

Mendel was exceptional in his talent of combining his skill in observation and his expertise in quantitative analysis. Mendel saw details that other missed, and saw through the noise of other artifacts so that he could classify his measurements and observations in a clearly formed and articulated manner. Then he took the data he collected and transformed it into conclusions which on the one hand explained what phenomenon was occurring but more importantly could be employed by others in not only duplicating his work but in expanding it to areas, animals, and plants that he may never even heard of. Mendel created an artifact, a factotum, the gene, which controlled color, shape, and other characteristics before one even imagined what that gene really was. The genius of Mendel was clarity of observation combined with consistency and competency of analysis.

The work reported in this book hopefully attempts to build upon that Mendellian paradigm. It did not start with theory, it started with working with plants from a specific Genus and asking the simple question, how did we get from a dozen or so simple species into a complex mass of hybrids? There were no new genes created, so there must be a process occurring where the genes are interacting in a complex fashion. How can one explain this phenomenon? I saw yellow H minor, reddish H aurantiaca, fragrant and long tubed H citrina, and the other species. I then saw the new hybrids, and grew my own new hybrids, and then asked; how did these complex patterns evolve?

Botany has been viewed as a science focused on collecting and classifying. The primary focus of the latest efforts of genetic engineering and analysis has all too frequently been focused on the areas of medical research where the breakthroughs can more immediately help mankind. However the daylily represents an interesting genus worth of study because it has just within the last century seen dramatic steps from a species isolated in hill sides of China, Japan, Korea, and western Russia, into an explosive genus of plants with strong horticultural interest. The genus has provided a Petri dish for understanding genetic pathways in a most visible manner, namely the complexities of flower color and patterning.

In an amazing paper published just before his death in 1954, Alan Turing, the brilliant scientist and mathematician responsible for the conception of the first true computer, the Turing Machine, Turing published a study of how plants and animals get color patterns. He did not yet understand the concept of a gene and he further had no understanding of color secondary pathways controlled by genes. He just postulated such a mechanism and then went and described how it would work. Thus from this he described Zebra stripes and coloration in plant leaves.

Ironically, just over the hill a piece in England at the same time Watson and Crick at Cambridge had developed a model for DNA functioning. And just a short distance to the

west, in Dublin, and a short time earlier, Erwin Schrödinger had developed and discussed the same set of issues, indeed the ones motivating Watson, in his small book called "What is Life". In the fifty years or more since this time the advancements have been monumental.

This book was the outgrowth of more than twenty years of hybridizing daylilies. Taking one parent and crossing to another led again and again to asking the question, how does this happen? I could ascribe the process to one color but not to patterns. The explosion of patterned daylilies as documented by Petit was another drive in that direction. What allowed these colors to explode in ever more advancing complexity? How could one model these, how could one predict what could result and how one could even engineer the result.

My approach to this study was as an engineer, also as one trained in medicine and botany. As one trained in botany and as one trained in medicine, one looks at the detail, the fine patterns that occur. When hybridizing, one gets to know every plant, every characteristic; the petals and sepals, the throats, the eyezones, and the slightest color patterning that may be there. As a botanist one also looks at the twelve species from which this explosion of diversity has arising and asked what in the genetic makeup allows this to occur. Then as an engineer, one asks why and how; how did this happen and how can I control this process. This is not mystery, it is just a set of natural processes and if we can understand them then we can control them.

Thus this book reflects this desire to understand and to control. This is not a book of pretty pictures for hybridizers. This book is a challenge for those who really want to make a change in this Genus and understand why. The classic hybridizers of daylilies have been 19th century geneticists at best. They used Darwinian artifacts and generally to not advantage. This book is an attempt to use this genus as a motivation and as a vehicle to explore many new and complex issues in genetics.

More than half of the material in this book is original and is not derivative. This includes the analysis of secondary pathways, the analysis of Turing tessellation, the details on species classification, and the use of backcrossing for the purpose of both hybridizing as well as validating genetic presence. The classic problem with the presentation of original material in book form is always the need to have second and third eyes clarifying the material via rewrites and clarification. Thus I apologize to the reader if the obviousness of some of the discussions may thus lend themselves to potential obfuscation; I have tried my best to rewrite most of what was original so as to bring it into conformity with a book presentation.

This book may appear to have a narrow focus but like any study on say *Caenorhabditis elegans,* **or** *Arabidopsis thaliana, or* other classic species, Hemerocallis presents a very graphic example of genes in action. Unlike *Arabidopsis* which is a small plant and has a shorter life cycle than does Hemerocallis, the Genus Hemerocallis has the feature that it

presents great color and pattern variation in the flower. It is a wonderful test-bed for the secondary pathway processes which control so many things in nature. This book thus presents a wide scope of ideas using a single vehicle. The ideas presented herein are wrapped in the paradigm of Hemerocallis but apply equally to all species including Homo sapiens. There are many disorders in the human which are secondary pathway disorders but for which we have not been able to remedy. Perhaps this vehicle of the Hemerocallis may assist the enlightening of some of these issues.

The book is an outgrowth of my efforts at MIT with my doctoral and post doctoral students but it is directed to the generally well educated individual interested in understanding the why and how of the process of hybridizing. Moreover the intent is to also allow this vehicle to be a stepping stone to applying the results presented in a more general manner.

Terrence P. McGarty
July 2009

Florham Park, NJ
Cambridge, MA

1 INTRODUCTION

The genus Hemerocallis is an intriguing example of diversity as well as example of the recent changes which man has made by selective breeding or hybridizing. The genus is primarily of East Asian origin, from China, Korea, Japan and eastern Russia, and is a hardy plant surviving from North American Zones 10 through 2. The genus, for the most part, is also fairly disease resistant and propagates very well. Two of the species in the genus are cold sensitive and survive only in southern regions.

However the most amazing thing about this genus is that it has been hybridized expansively only for the past one hundred years and many records associated with that hybridizing have been kept and are available, on line.

There are almost two dozen species recognized in this genus. We shall focus on twelve which are grown here in North America and which are winter hardy. The two non-winter hardy members of the genus we will avoid because of their limited availability. Understanding and identifying species is a science and an art. It is built upon the ability to see fine detail and to describe the detail in a consistent manner.

1.1 HYBRIDIZING

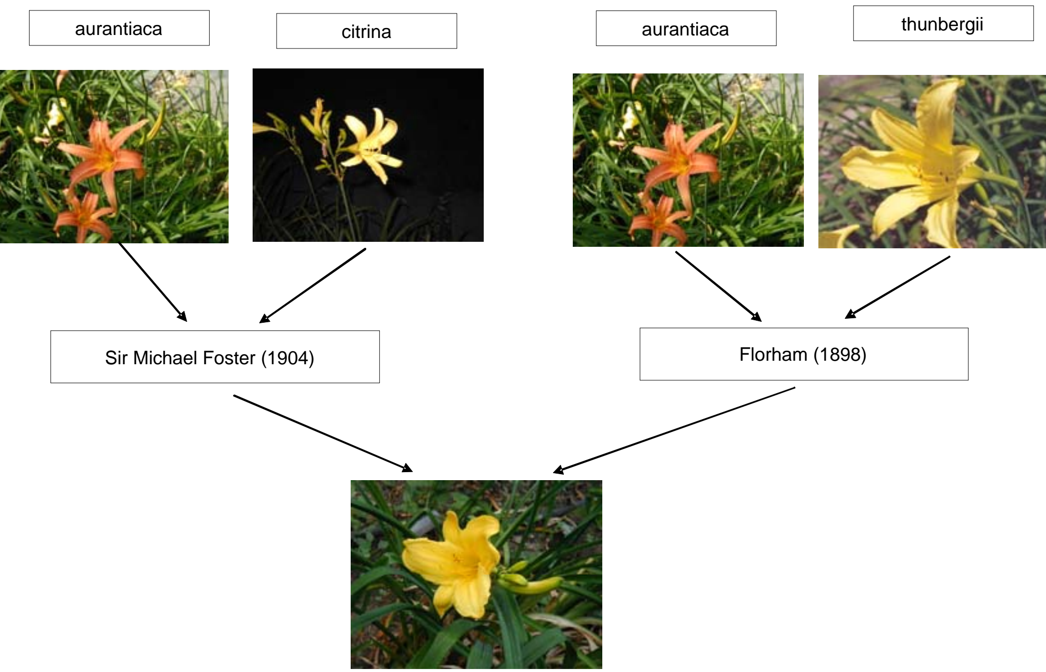
Hybridizing is the name for breeding of Hemerocallis. There is a slight twist on terminology. The use of the term hybrids and hybridizing and breed and breeding are often confused and confusing[1]. Hybrids are frequently defined as crosses between two parent having significant differences in genetic makeup. This may mean crossing what we would call two species. Breeding is the process of taking plants in the same species and crossing them to seek specific characteristics. Thus in Hemerocallis the initial crosses was true hybrids. Taking H minor and H aurantiaca would yield a hybrid. However in the current time of hybridizing the hybridizers all too often are really inbreeding, taking plants with the same or similar extreme characteristics such as patterns and breeding them to intensify the pattern. To keep with current colloquial usage we shall maintain the application of the term hybridizing.

Hybridizers fall into several categories; the professional, the professional amateur, and the amateur. By using these terms I mean no demeaning of any one, it is style and not quality that is attributed to each. The professional is generally one with a strong academic background who keeps meticulous records and is seeking to understand how the species has changed and how the colors and variations in form arise. Stout was the prototypical example of the professional.

[1] See Oxford Dictionary of Botany, 192, Oxford. The definitions we provide here are from that source.

The hybridizing of Hemerocallis has been documented by several other authors and will be not be detailed herein.[2] However we do want to present an overview of the changes which have occurred as a basis for the questions which have been presented earlier.

Let us consider an early hybrid, the plant Hyperion. This is a strong yellow flower and it is still for sale as of today. The Hyperion hybrid cross is shown below:



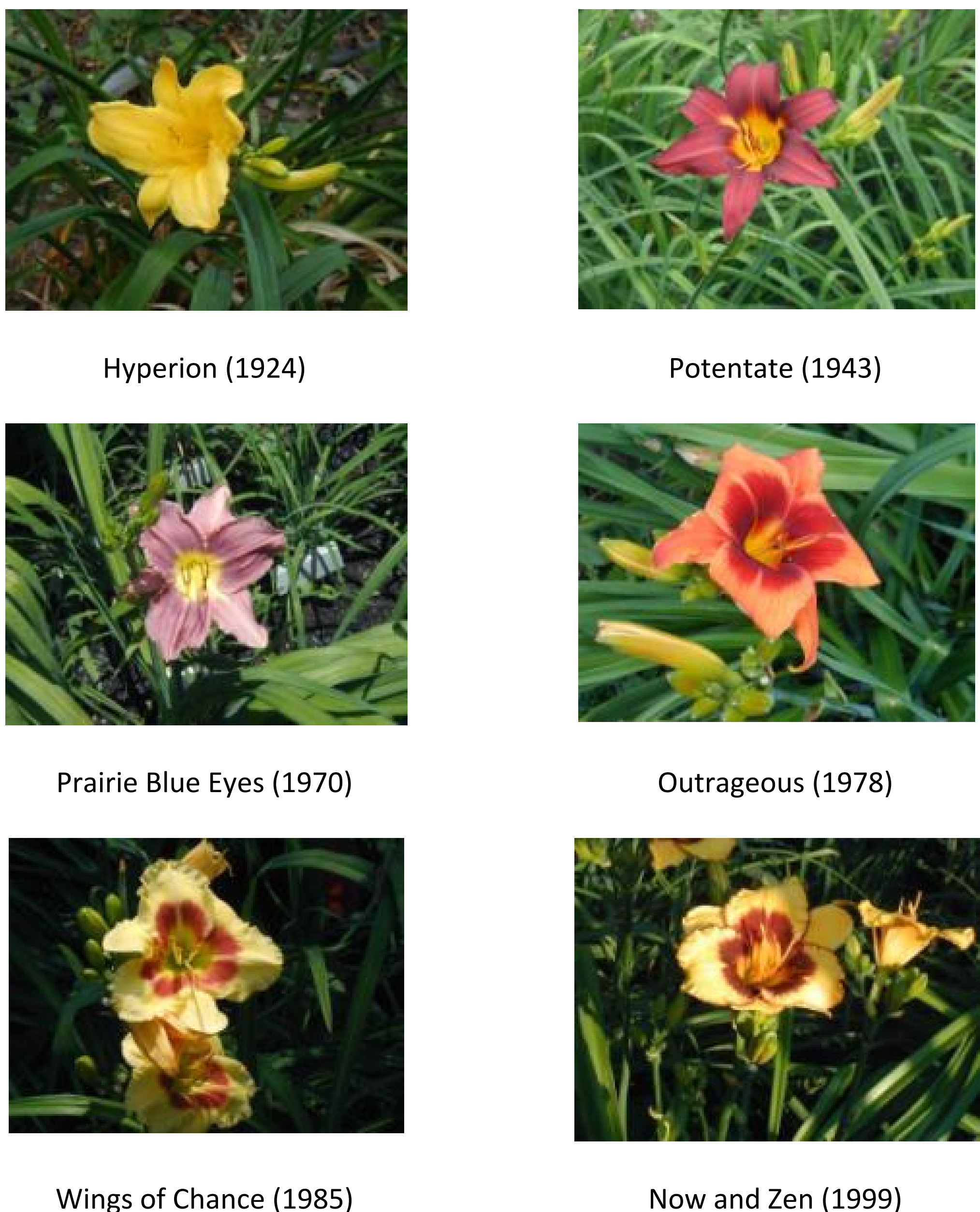
We note that in the Hyperion cross we have H aurantiaca, a reddish parent, on both sides of the plant. Also we have a nocturnal H citrina and a closely related species to citrina the H thunbergerii. We call the parent generation the F0 generation and the first set of offspring the F1. Thus we can call Hyperion an F2 generation from an all species set of parents. The F2 result is thus the all yellow Hyperion. Although we have no copies of the parents Florham and Sir Michael Foster in our collection we are led to believe by records that they are also yellow[3].

---

[2] There are many good works on the hybrids. The classic is Munson albeit a bit dated. The recent work by Peat and Petit is excellent.

[3] See Stout, pp 48-49 for Florham which is stated to be canary yellow. Also p 71 for Sir Michael Foster declared to be a clear yellow. This may imply that red is recessive, and that yellow is dominant. Yet there is not adequate controlled set of data to validate this.

The following Table depicts seventy five years of hybridizing from Hyperion in 1924 through Now and Zen 1999. Potentate was one of the first truly red flowers and became a benchmark which holds even today. Prairie Blue Eyes was an attempt to obtain a blue, close but not totally there. There have been attempts at whites as well, with considerable success achieved in the white hybrids. The other more recent flowers show increasing complexity. The 1999 flower, Now and Zen, shows an eyezone, a colored or tinted edge to the sepals and petals and a well demarcated throat region as well.



Hyperion (1924) | Potentate (1943)

Prairie Blue Eyes (1970) | Outrageous (1978)

Wings of Chance (1985) | Now and Zen (1999)

With the above development there are several questions which we can ask:

1. Hyperion was the F2 development of species crosses, or at least that is the way it appears from the early literature. Hyperion has a substantially different form from any of the F1 plants or the species which are both yellow, despite the F0 parents which were both reddish. What accounts for this change? Hyperion has a color which is a stronger yellow than citrina and does not reflect any of the variegation of the other F0 parents. What set of genes have been suppressed?
2. Potentate has what is called a "throat" a gold region inside the flower. Throats like this do appear in the species. What controls the throat characteristic? The transition between throat color and the predominant color of the flower is very abrupt, what genetic switch allows for this abrupt change?
3. Prairie Blue Eyes is one of the early attempts to get a blue. One would assume it would be possible. Only recently has genetic engineering produced a blue rose. The question then is can we breed a blue daylily or does it require some form of genetic engineering, that is introducing a gene not naturally present in the daylily?
4. The Outrageous flower starts to show the dramatic change in form as well as color. It is a recurved flower with significant color variation. In this case form as well as color is being changed. What are the genetic linkages between them?
5. Wings of Chance and Now and Zen show how quickly genetic variation can proceed. The throat becomes an eyezone, a region from yellow to red to yellow. And in Now and Zen we see edging colors appear on the end of the petals and sepals. Again, what is the gene expression control mechanism which effects this unstable change? We have many examples from dynamic systems, can we apply them here?

What we then ask is;

(i) knowing the species and assuming the species have some definable set of genes, and

(ii) that the genes in the species express themselves so as to generate the colors we see on the species, and

(iii) furthermore given that we have not introduced any new genes into the new hybrids nor have we mutated any of these plants,

(iv) how, through hybridizing alone, have we managed to allow new combinations of existing genes to be expressed and to have existing genes expressed at new and greater or lesser rates than the species plants?

Simply, the question which drives the progress in this book is:

Given a set of stable species which reproduce naturally in nature, and given that the genes which control these species are the totality of all genes available, how is it possible to attain so complex and variegated a set of hybrids in so short a period of time?

If we look at the genus Hemerocallis, we see that we have been obtaining hybrids for just over a hundred years. If we assume that we have had 5,000 hybridizers over that period, and if we assume that it takes three years to assess a good hybrid, and if we assume we now have 50,000 registered plants and we assume we select one out of a hundred crosses we can make a simple estimate as follows. There have been approximately 5 million crosses over 100 years, albeit weighted heavily toward the current time. Hyperion was an F2 cross. The recent introductions are thus F30 to F40 crosses. Thus we have rearranged by a process of "intelligent design" hybrids which reflect a certain set of characteristics. We have been able to introduce patterns, eyezones, edging, bi-colors and they like in just a few dozen generations. "Natural selection" in the Darwinian world has done just the opposite, it has selected a dozen or so species which remain stable and reproduce effectively in their own worlds.

1.2 A Platform for Discovery

Hemerocallis is also an interesting platform for discovery. On the one hand, the extensive hybridizing allows for the analysis of how colors and patterns evolve in plat flowers. The other variables such as leaf morphology and root morphology seem to have varied much less than flower color and form. People have been able to introduce extreme variants in flower formation and color by the extensive hybridizing efforts. Once a unique aspect is obtained then the hybridizers home in on that characteristic and breed as strongly as possible to both retain it as well as seek variation on the same theme. Thus we see many spiders and other similar variants, double plants, multi petals, and various types of coloration. In less than a hundred years of such hybridizing the variation in color and form has become so extensive that one looks at the species and can barely recognize that a mere hundred years ago that that was all there was available.

The development of color in hybrids of plants can be viewed in many ways. One, the Mendellian approach, is that there is some element, called a gene, which is on a chromosome, and there is some mixing set of rules, dominant and recessive, which when applied allow for the control of the color development. The recent point of view is that there are genetic pathways which are controlled by enzymes, proteins, and that understanding the mechanisms of the control of these pathways is key to understanding the process.

However, the questions we raise herein, and then seek answers in the current literature, and finally propose possible paths of inquiry are as follows:

1. Given a dozen or more species plants which are relatively stable and consistent in the wild, how does the variation in color in hybrids arise? Namely, what is the cellular basis of color, and moreover what is the genetic set of mechanisms which control this.

2. Given the complexity of color, form and variegation in the hybrids, what is the genetic basis for the control mechanisms intracell and intercell? For example, how such colorations as eyezones formed and what are is the intercellular communications mechanisms which affect this.

3. Given what now appears to be a set of well understood pathways control mechanisms by enzymes produced within the cell and the gene control mechanisms for expression of these proteins, how are these combined to produce intra cellular coloration and what are the inter cellular communications which spread the colors out over the inflorescence.

4. Given that we can answer the above, can we generate a mathematical control modem for gene expression and control and using the model approach the coloration problem as a problem of system identification or inversion.

5. Given that we could solve the above problem, then how could we invert the inversion and apply positive control to coloration and produce whatever color and for we would so desire.

We attempt in this Book to address these issues and set forth a combined understanding of what appears to be at this time a fragmented set of research efforts.

Our approach in this Book is fairly straightforward. We focus on a specific genus, Hemerocallis, and on a specific part of the plant in that genus, the inflorescence. The questions we ask are; (i) what is the cause of the colors we see in the flowers given what was in the original species, (ii) what are then pathways that generate the substances which produce the colors and what enzymes control the pathways, (iii) how can we develop a system level model for this process, (iv) can we, using the system model, develop methods to develop desired colors.

One of the first questions which can be asked is why this genus? There are many reasons for using this genus to stud the process of gene expression. The following are a few reasons:

1. The genus has been hybridized for just the last one hundred years. Thus there is a wealth of hybridization cross information to be able to assess what the genetic makeup is of the novel hybrids.
2. A great deal of recent research has provided detailed explanations for the control of color pathways and these apply directly to the genus.
3. The hybrids have been able to express color and form variants which are quite striking and allow for a clear identification of both pathways and gene expression mechanisms.

4. The genus is composed of s finite set of stable species. The underlying species of the genus Hemerocallis is generally well circumscribed and is currently under extensive study.
5. The genus does not appear to have significant transposon effects or viral effects. Unlike tulips and others species where viral changes are the generally more reflective cause of phenotypic change or in corn where transposons have a significant impact on phenotype, Hemerocallis appears to e dominated by gene expression changes.
6. The genus has multiple hybridizers making multiple changes per years. The American Hemerocallis Society lists over 50,000 hybrids and there are well over 500 active hybridizers in the US alone.[4] A typical hybridizer may make anywhere from 200 to 5,000 crosses per year and keep 1% of the crosses for registration, the remaining 99% going into a possible general pool of hybridizers "road kill".

For these and many other reasons Hemerocallis is an attractive genus.

1.3 OUTLINE OF BOOK

The book provides an overview of the genus and also provides details regarding its classification, hybridizing and basic biological processes as well as it growing characteristics. The book also investigates various genetic elements of the genus and in particular looks at the issue of color in the inflorescences. Prior work on this genus has been highly speculative and dated in terms of the current state of the art in genetic development.

The book contains the following elements:

1.3.1 FORM AND SPECIES

First we look at the issue of form. The process of classification is a process of specifying differences in form as observed in many dimensions; the shape of the flower, the shape and complexity of the stems, the structure of the roots. Also included is the temporal nature of the plant, when does it bloom, and how does it deal with environmental conditions. We frequently want to characterize the genus into species and possible even in lower orders of classifications. To do this we require the ability to differentiate one from the other and it is this collection of differing forms which allows us to take that path.

[4] The author is one of those hybridizers have introduced over fifty cultivars in the past twenty years. The author believes that it is essential in any science that one must have hands on experience with the subject matter at hand, either in the micro or macro, of optimally, both.

Developing names and dividing the plant into pats is the classic approach to characterizing the genus. In fact it becomes the basic requirement of any and all plant classifications for the past thousands of years, especially since Linnaeus. We begin with the plant and then spend time defining and describing its parts and what to look for in analyzing one plant from another. One of the things that anyone who works with plants, or frankly any species, has to do is to become familiar with the many micro details of the plants. One must look closely, then look closer, and then stand back and do it all over again. Thus when we provide for a detailed set of descriptors for a plant we do so initially to understand it structure, then to understand its differences between species and then again to obtain metrics with which we can compare one species to another.

The next step in any classification processes it to develop, define, and qualify the metrics which we will use in the plant classification process. Thus the leaf length, the leaf width, the flower color, the date of first bloom, the number of flowers per scape, the size of the bloom, and the like are all elements. Our approach herein is to develop quantitative metrics for classifying. We avoid the more classic use of word descriptors since we focus on the observables which are measurable.

Then we ask the question; what is a species? This question has been asked for thousands of years, dating back at least to Aristotle. Is nothing more than a method of classifying similarities or differences or is it a biological barrier. Ernst Mayr, the Harvard biologist, said species are entities which can inter-propagate, that is two plants of the same species can cross pollinate. If they cannot then they are of a different species. Is a German a different Species of the genus Homo from a Thai? We would think not under the rule of Mayr, for indeed a German and a Thai may have offspring. We thus think of Homo sapiens as that group which can propagate amongst itself. Yet with Hemerocallis, all of the species can cross pollinate, almost. Thus under the Mayr rule they are all one species.

The first step is to review the genus. Hemerocallis has about a dozen species, most of which can be interbred with one another. Some are self sterile and many can be bred. Hemerocallis has been hybridized for the past hundred years and many records of their ancestry exist. Thus it impossible uses this genus to track many of the genetic linkages. The twelve species are all consistent within the species, there is some local variation and some geographical variation but it is possible to develop classification keys which generally are predictable and stable. The different species have a similar form but the colors vary between species, and there is even some color variation within species. In addition some of the current phenotypic species may also be variants of another species. We do not get into these arguments since the ultimate determinator will be a genetic classification, much of which is already under way. We then proceed to show how in the one hindered years we have been able to introduce significant variation in this genus. We then use this as a basis for developing a discussion as to why and how can it is controlled.

Then we provide an overview of the species. We focus on the twelve that we have been growing for the past twenty years and provide details on the others which we have not had direct experience with. The overview of the species looks in details at the differences in forms between flowers, scapes, leaves, pods, seeds, roots and also the temporal a cultural characteristic of them. One of the things we find amongst the species is intra species variability as well as inter-species variability. There are certain species which will crowd out other species, and some which bloom quite early, in April in our zone and others which bloom quite late, October in our zone. We see H middendorfii in April and we see H fulva sempervirens in October. We see H minor in late April and again in late September.

1.3.2 Genes and Gene Expression

We then present a detailed analysis of the genetic issues which will be used in better understanding the genus. This is an overview of modern genetics and also looks into the issue of gene expression and modulation. This chapter is not a classic genetics study but looks at the genes from the DNA level.

We then provide a details overview of cell genetics and how activators and repressors are key elements in the overall expression of enzymes and in turn the development of color. We present a review of the cell elements and especially the process of gene expression. We discuss activators and repressors and the mechanisms of their actions. Their existence results from the work of Monod and Jacob in the early 1960s.

This chapter also looks in detail at the secondary pathways which are players of critical roles in the lives of plants and all living organisms. The plant secondary pathways we focus on are those controlling plant colors. The typical pathways are those for anthocyanins, carotenoids and the like. These pathways are controlled by proteins generated by genes. The proteins are catalysts which control the pathway, making it go faster or slower. We analyze and then model these processes.

1.3.3 Phylogenetics and Classification

We then discuss the problem of classification. We have presented the various historical species in the Genus and we have also discussed the genetic issues related to the genus, focusing on the coloration issues. We then address the issues related to classifying. We commence the chapter with a discussion of what a classification must accomplish and then look at the various methods which have and are currently employed. Then we look at the genetic methods of classification and this leads to a study of the ways to test for the presence of certain genes, their artifacts and segments relating to known genes. This is a discussion of the currently available techniques and they will become critical in our latter analysis and synthesis efforts.

Then we proceed to provide further detail on the classifying methods which can apply to this set of species. We accomplishes this in some details showing the use of the data one can collect using the methods described and how effective each of the classifiers are.

Finally we look at how speciation may occur at the gene level and in so doing we look at the imputed speciation which may have occurred in Hemerocallis.

1.3.4 FLOWER COLOR AND COLOR EXPRESSION

We then proceed to present an overview of the process of developing color in flowers. We present an overview of the anthocyanins, flavonols, and carotenoids. We review their pathways and summarize recent research which had identified the enzymes on each link of the pathway and the genes controlling those enzymes. This has been accomplished over the past few years and is critical to the understanding of the overall system approach.

The color of a flower and subsequently the issue of its patterning are examined in this chapter. We first examine the issue of color as observed by humans and then how it is characterized and measured. The phenomenon of color measurement and characterization is a key element in performing any validation of the theories developed. We look at the issue of color.

We then apply these models to the secondary pathways we have discussed earlier.

There has been a recent development in the biological community of applying system models to biological systems. We build on that effort and develop medals for the expression of flower colors. Simply put, we recognize that color is a result of a mixture of secondary plant products such as anthocyanins. We can from the color of a flower determine what the mixture of each anthocyanin is. The concentration of an anthocyanin is a result of the concentration of the enzymes in the pathway which produces the anthocyanin, and typically the lowest enzyme concentration is the dominant factor. We also know that the concentrations of the enzymes are a result of activators and repressors, proteins also generated in a cell, which turn on or turn off the enzyme controlling the pathway.

Combining these ideas we can develop a top down system model for color. The output or observation equation is the color, and the system equation is a dynamic process wherein the states are the protein concentrations from a large enough set of gene expressions, wherein genes are allowed to control other genes via an nth order dynamic process. We also allow for uncertainty by adding a “noise” process which converts the overall system model into a linear dynamic stochastic system with observables. We then extend that model from a single cell to a matrix of interconnected cells. This then allows us to explore the processes one sees in the development of eyezones and other sharp

transitions of color in flowers. We use models which have been previously studied for color variation and apply those to the flower.

1.3.5 Flower Color and Patterning

This Chapter discusses the issue of pattering in flowers. This is also a form of tessellation or coloring. The specific model which we employ is the Turing model. We use the recently discussed patternings of Petit and we demonstrate that the Turing model can be the descriptor of how these patterns are affected. This analysis is a joining of the genetic models, the secondary pathway models, and the color effects in an evolving hybrid.

In this Chapter we develop the Turing model, we then relate it to the secondary pathway model we have validated, and then we apply it to the Petit patterns. We then take sample from known species and hybrids and look at them on the cellular model to validate the Turing method and demonstrate it viability. This then allows us to use this model to predict how certain patterns may evolve and be controlled at the genetic level.

1.3.6 Classic Genetics and Hybridizing

In this chapter we return to the classic Mendel model of genetic hybridizing. We review the classic Mendellian genetic approach and attempt to apply it to the species. We see that the concept of a gene as used by Mendel falls apart quickly in Hemerocallis and its hybrids. One would have to posit thousands of genes just for color and form and variegation, not to mention the other factors. The Mendel approach may work well for peas with limited characteristics but it has no places in this analysis.

In the early 1970s, Joanne Norton had published several articles on Hemerocallis genetics in the Journal of the American Hemerocallis Society. Norton had a PhD in the field of agriculture and had apparently done doctoral work in classic genetics. At this time however there was a massive explosion in the understanding of DNA, none of which is reflected in her work. She established the state of the art in Hemerocallis for the past forty years. She used a Mendellian approach but with no adequate data to back up her assertions. She postulated results and then based them upon anecdotal data at best. However it was a first attempt. In this Chapter we do not attempt to reproduce her work but we do use the Mendellian approach. This is used to discuss several specific methods used in classic breeding and hybridizing methods. This is then used to clarify the many approaches used by hybridizers.

Hybridizing as currently practiced is an art and not a science. Frequently a small community of hybridizers share genetic material of the current version of what is fashionable. Thus there are those sharing spider like plants and those with eye zones or patterns. Then they inbreed the characteristic as many times as they can.

We end this Chapter with a detailed analysis of backcrossing, a way to drive a specific gene into an existing pure bred. We assume we want a red H citrina, thus we use back crossing to achieve that goal. We also show how backcrossing can also yield great insight to the genetic structure.

1.3.7 HYBRIDIZERS

The hybridizers of this Genus have been a strong community of amateurs for the most part. They are not agricultural experts as one may find in hybridizing corn or rice, but of the type one may find in many types of ornamental plants. They fall into mat types which seem to defy classification. At one end are the Stout and Apps camp, both of whom have PhDs and are educated and accomplished botanists in their own rights. Stout spent decades at the New York Botanical Garden hybridizing many of the early introductions. Apps, who has just retired at the time this Book is written, has recognized the difference between a display plant and a horticultural plant, namely one grown to be grown, not just for display. Then there are monks, retired school teachers and small farmers, and artists, and everything in between. Each tries to set themselves apart by their apparent uniqueness. There are eyed plants, spiders, patterned, edged, and every variation known to man.

2 FORM AND SPECIES

This Chapter introduces the several species of the Genus Hemerocallis. We start with discussions of the various forms of the plant that the species have and provide a basis for establishing the different species. We look at the flowers, buds, pods, seeds, scapes, branches, leaves and roots. The first thing that a botanist, and any scientists, engineer, or physician must do is to learn how to "look" with all the senses. This would include the sense of smell, taste, touch, not just sight. The texture of the leaf, of the root, of the petal will be important differentiator between species. The scent of the flower is important, for example, H citrina is quite fragrant and this is an attractive characteristic which we would like to preserve.

Then we present an overview of the species. We examine the details of all of the elements we have discussed in form for each of the species. Again the broad concept of "looking" must be evoked. We present a detailed comparison of each of the species and examine how they have been characterized by others.

The approach taken in this Chapter is to look at the species from many viewpoints. The goal is to demonstrate that observations are critical, and that one must continue to improve the detail in observation time after time. The minutest detail will become important. This is not just a listing of the species but an excursion through the way one may learn what they are and how they differ and how they have similar characteristics. Thus the reader must view this chapter as just such a journey, the Appendix listing the species and their detail but in between we look at parts and species and species and parts. We always come back to detail.

2.1 GENOTYPE AND PHENOTYPE

Phenotypes are what we see, smell, hear, touch, taste; they are the interactions between some creatures, in our case a plant, which we may use to identify the plant. In the genus Hemerocallis, the phenotype may be the color, color patterns, size, time of bloom, odor, texture of the flower, and other definable characteristics that we see when we observe the plant.

Genotype is what the gene has as specific content, its specific DNA. The production of a phenotype is frequently driven by the expression of a gene. The gene "expresses" itself in a very special manner. The DNA is wrapped in tight coils.

The model we will build upon appears as in the Figure below. This is the canonical model for gene expression. We assume that there is some collection of secondary pathways, and that these pathways result in chemical products that are directly related to a phenotype; a darker red flower, a longer leaf, a taller scape. That these pathways are

modulated in some manner by proteins generated from within a cell. That the proteins are the result of some entity called a gene. That the gene can be an assembly of bases and the gene may itself be modulated up or down by activator or repressor proteins respectively generated by other genes or even the same gene. Thus we model the cell as a dynamic system and further we argue that this system has certain random elements which we shall include latter.

It is the output of this genetic process that we get the plant in its full temporal and spatial existence. The above model of the gene is one in which we see the beginnings of some form of feedback. We see the activator and repressor genes as the basis for this element. However this may be expanded even further, we show this below. Note we show that the Gene K can be influenced by other Genes, as well as the products of the pathways as well as by the environment. The Environment can modulate the pathway which by being fed back to another controlling gene can then modulate the activating gene. This process is a complex process and exceeds what we would have imagined from the simple Mendellian gene theory.

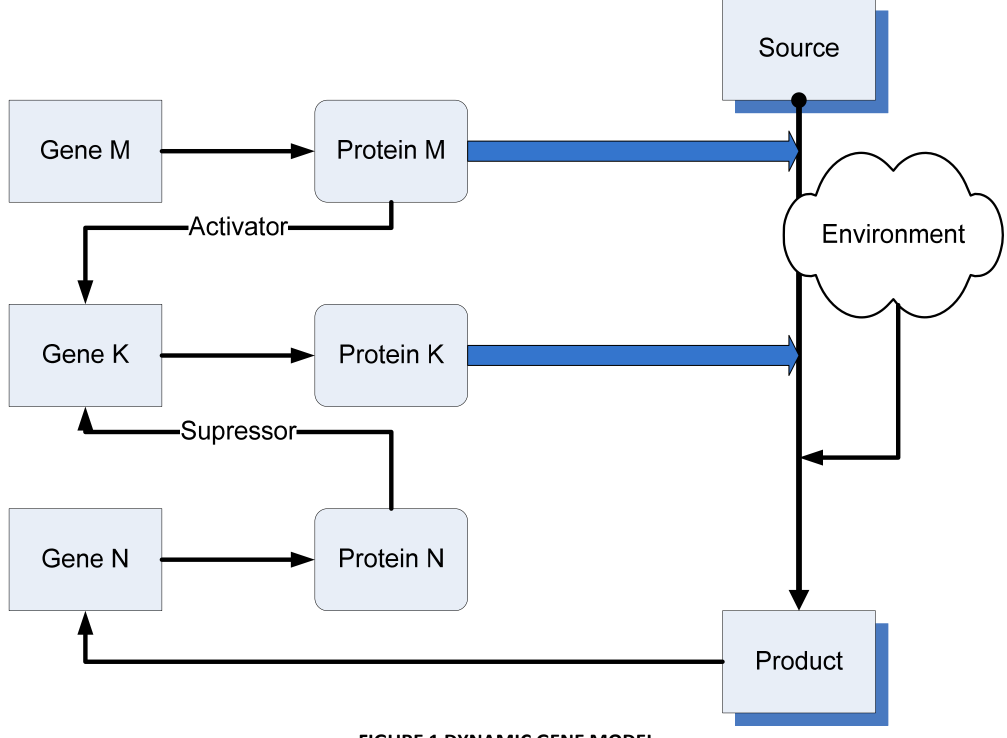
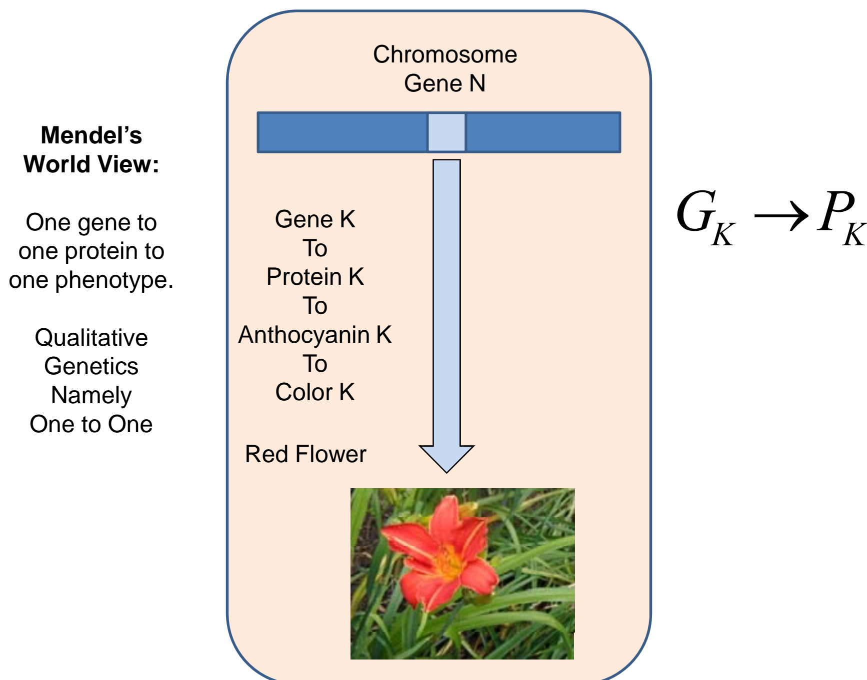


FIGURE 1 DYNAMIC GENE MODEL

Now back one again to the Mendellian Gene model. Although Mendel and his model was not so rigidly simple, for he did admit some other influences as well as variation, we will call the simple Gene and Phenotype combination the Mendel Model. Namely in this model we assume the existence of a Gene and then we further assume that there is some phenotypic characteristic such as flower color which maps one to one onto this gene. One gene and one phenotypic character. The phenotypic characters further have countable and discrete values. The flower is red, yellow, and green. There are no blends

and there are a limited numbers. Then there is a gene for red, a gene for yellow and a gene for green. The gene is at the same place on the chromosome and the gene just somehow changes to produce a different color. In addition the genes are dominant in some order. That is if there is one red gene, of the two on the chromosome, then we get red, if not a red but a yellow we get yellow, and we get green if and only if there are all green genes, namely two.



Now there is a second model, based upon our understanding of DNA and the Watson Crick world. However this model goes well beyond the simple Watson Crick model. Here we assume we have long segments of DNA with many exons and many more introns. The gene as we know it is the result of the cellular processes which assemble the exons into a block of DNA which RNA will use to in turn generate a protein. In reality what happens is that the exons may be recombined to generate RNA in a variety of fashions. The result of that process, as well as the dynamic model we depicted above is that the phenotypic characteristic, say leaf length or width, or date of first bloom, takes on the character of a random variable. It has a set of values whose probability distribution may be of some form. We use as an example a standard Gaussian curve.

The following Table depicts Morphological traits (mean and standard deviation) of *Hemerocallis citrina* and *Hemerocallis fulva*, their F1 hybrids, and individuals in the hybrid population the standard deviation is given in parentheses

	H fulva	H citrina	F1
No. of scapes	72	74	55
Flower tube length (mm)	32.40	45.28	32.27
Petal length (mm)	92.51	81.13	78.60
Petal width (mm)	15.42	14.50	16.01
Stamen length (mm)	77.02	68.04	64.39
Pistil length (mm)	97.34	76.19	77.73

2.2 SPECIES

The development of the concept of a species has been long and fraught with controversy. A species was initially a construct which allowed for finer and fine classification of plants, animals, and even minerals. It was to Linnaeus merely a means to classify. The work by Mayr and many others gave to this concept much greater meaning. It is the Mayr meaning, as modified by many others, which we will use herein.

Let us start with the three properties which Mayr required of a species. These are[5]:

(1) The members of a species must be a reproductive unit. They must be able to mate and moreover they must mate.

(2) The species must be an ecological unit and interact with the other species in the environment. Thus the species must interact with not only other plants but other animals as well.

(3) The species is a genetic unit consisting of a large intercommunicating gene pool. The species is not just one plant or even just a handful. The genes in the species must have some diversity to them.

To give an example from the Hemerocallis we consider the work of Hasegawa et al on the interaction between H fulva and H citrina[6]. The authors considered these two species and their growing habits. H fulva and H citrina are both only open and available for pollination for about 12 hours a day. H fulva blooms in early morning and till late afternoon and H citrina from late afternoon till early morning. Their bloom times rarely overlap. In addition H fulva is a colorful flower and has no scent and is pollinated by swallowtail butterflies, whereas H citrina has great scent and is a pale yellow and is

[5] See de Queiroz, NAS p. 6605.

[6] See Hasegawa et al Bot Soc Japan 2005.

pollinated by nocturnal hawkmoths. Thus they live in the same environment but are ecologically isolated. The two share the same space, and interbreed at rare occasions.

This example by Hasegawa represents a typical example of two species as would be understood by Mayr. The H fulva and H citrina each self pollinate and grow, they live within their ecological environment, and it has been shown that within each species there is significant intra species genetic diversity. Mayr's concept of a species is a biological construct, as shown in the preceding example. It is not just a concept driven by form, shape, color and the like. It is a concept involved in the very existence and environment of the plant, of the species. It means it is a living group which has found a way to associate and co-habit its environment, in turn taking on characteristics which have great outward similarity.

For the Mayr view, therefore, the species is analogous to the members of the other fundamental biological organizations such as the cell. The species is a biological or living construct which has boundaries defined by its very existence and interaction with its environment. To Mayr, species are groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups. This is the "biological species" concept. This set of constructs, this definition in the context of a living environment and ecology, this definition in terms of what we would understand as say a cell, will make the definition of species in the genus Hemerocallis a complex issue. In our above example, we see that H fulva and H citrina have that isolation, isolated not geographically but reproductively, and isolated reproductively by time.

The following Table is from de Queiros, p 603 and depicts a summary of the differing species concepts. The Table depicts Properties, in addition to existence as a separately evolving metapopulation lineage, commonly treated as necessary properties of species

Species Concept or Definition	*Property*
Biological species concept definition (isolation species concept)	Potential interbreeding (intrinsic reproductive isolation)
Recognition species concept	Shared specific mate recognition or fertilization system
Ecological species concept	Same niche or adaptive zone
Monophyly version of the phylogenetic species concept Genealogical species concept	Monophyly (as inferred from apomorphy or exclusive coalescence of gene trees)
Phenetic species concept	Form a phenetic cluster (quantitative difference)
Diagnosable version of the phylogenetic species concept or some interpretations of the evolutionary species concept	Form a diagnosable group (fixed qualitative difference)
Genotypic cluster species definition	Form a genotypic cluster

Hybrids of the two species are found, and they are viewed as just that, hybrids. They are crosses between species, and consistent with the true definition of a hybrid, that is what they become.

The Mayr construct then leads to the question of looking at say H citrina at many different points, and asking what is common amongst them that make a disparate geographical group all the same species, when there is a geographically proximate group which is of two species. What is the isolating factor that separates and what is the ecological or biological factor which combines?

Species are meta-populations, which are fundamentally different from the taxa above and below them. A genus is a somewhat arbitrary classification containing many species with some level of commonality. A variety is some subset of a species which has certain phenotypic characteristics which make it somewhat special. Yet a species is a metapopulatine lineage that has a form of evolutionary status which causes a separation amongst and between them. One tries to define the boundaries between species in an objective manner.

Thus for our purposes, we shall use the Mayr construct, that they can breed within the group, that they have some form of ecological isolation, that there is some genetic intra-species variability, and that blatant inter-species breeding is not a stable result. We do not take the pure doctrine that species can only breed with species. We also have to deal with the definition of hybrids as crosses between species. This is an important issue. For all of the breeding in daylilies is called hybridization and it is just that term which can come under doubt.

2.3 Breeding and Hybridizing

Hybridizing has generally been thought to entail the genetic crossing of different species. Breeding is generally construed to mean the genetic crossing of the same species. Thus one may ask are the people who genetically cross daylilies hybridizers or breeders. Are these flowers that are produced hybrids as one may generally understand the term or the result of a breeding process?

How does the genetic material get transferred, by pollination? One may have natural pollination by insects, animals, the wind or human pollination which is the result of some selective process, say intelligent design, by a human.

The next issue we will look at is the influence on genes and color. The following two corn cobs show one of these influences, namely the jumping genes. This will be examined for this Genus.

FIGURE 2 CORN VARIEGATION

FIGURE 3 DETAILS OF CORN GENE EXPRESSION

2.4 Classification

To understand any plant Family or Genus, it has been classically required to have a detailed grasp of the form of the plant and the ability to differentiate one form from another. Thus, for an example, consider a plant with short and wide leaves versus one with long and narrow leaves. Another would be to consider a plant with yellow flowers

and those with red flowers. The list may continue. But being able to characterize and to articulate the many types of forms then gives us the ability to develop ways to classify the plants. This is the old style of classification. Of course, we will also look at the new methods of classifying through DNA, but that is for a latter Chapter.

What we intend to accomplish in this chapter are several things.

1. Define: To define what we mean by the different parts of the plant. This is a naming and characterization effort. Our intent is not to discuss detailed plant anatomy but to deal with Hemerocallis and to lay out the elements of the genus as is commonly available in the species.

2. Characterize: To then take those elements that are characterized and to apply to them specific means of stating them. Thus we can look at leaf height and width, and number of branches, number of buds per branch, and the like. There is also a less quantitative set of elements as relates for example to roots. We could look at roots and say they are long and cylindrical or bulbous. The quantitative approach is to say that they are x cm long and have a maximum diameter of y cm. The terms cylindrical and bulbous characterize what they look like to most people. The details of length and min and max diameters provide a quantitative statement of the same.

3. Classify: Then we proceed to methods of classifying the species. By focusing on Hemerocallis we have a simple genus to deal with but the classification problem is still a complex one. We can take all the measurements we have obtained from above and then we can ask the question as to where the boundary is between different species and what are the elements we should use to seek those boundaries. Is it leaf length, flower size, color, and the like? Classification is both science and art. In this Chapter we do classification with what is seen and observed, the phenotype, in latter Chapters we go to the genotype that is classifying by genetic analysis.

2.5 FORM

The plants we are investigating have several characters which will allow us to determine what species each of them is. Species is a way to separate plants from one another and there is a considerable debate in the botanical community as to what truly constitutes a species but more of that latter. In this Chapter we present a simple first order analysis of the plants which we are studying and establish a base for further analysis.

Our objectives are as follows:

- Provide a simple overview of the parts of the plant in the genus Hemerocallis so that they can be used in identification, differentiation and characterization.

-
- Provide a baseline for the different types of characters and demonstrate what some of the extremes are amongst the different members of the genus. This is not a detailed analysis of the different species; it merely presents an overview of the types of differences prefatory to a more detailed analysis.
-
- Establish a framework for the detailed discussion of characteristics and their classification so that a detailed analysis of species may be developed. Unlike many simpler presentations of Hemerocallis, the intent here is to establish a more rigorous set of structures so that a fuller understanding of what the separate species truly are.
-

2.5.1 Parts of the Plant

Before beginning we will provide a basic overview of the plant. We shall return to this latter for more detail. Simply, the plant is divided into several parts; roots, leaves, flowers, and the temporal elements of the flowers, buds and pods. These five elements will be all that we will focus on in looking at the phenotypes of the genus. The phenotypic characteristics, those that we see and use to distinguish one from another, will be employed subsequently in ascertaining the species and will allow us to develop a methodology to identify them.

However, as we shall see, the classification by phenotypic characteristics, the classic means of creating taxonomy, is under challenge, from genetic methods. These we shall use latter in our discussions. For the person looking at the plant in the wild, the phenotypic elements are the first things that can be viewed and used to identify the plant, not necessarily classify the plant. The use of standard terms will also be useful in communicating the plant in word form from one person to another. However, we would argue that the use of the words were essential in the world where pictures were costly and rare. In the world where pictures are much more common and become an essential part of any field work, aligning words and pictures becomes essential. Thus cylindrical versus bulbous roots will become much more evident when one looks at the roots of H. citrina and those of H. fulva. Then one may ask why the roots are different and the “why” connotes both an ecological and genetic essence. The posing of the important “why” will be a thread that we will try to weave throughout.

The plant parts we will look at are as follows:

1. Roots: These are the below ground portions of the plant used for water transfer and nutrient storage and transfer. They are also part of the mechanical and vegetative propagation part of the plant.

2. Shoots: These are the above ground portions of the plant and are composed of two major parts:

2.1 Leaves: These are the nutrient generating elements where the chlorophyll is located and where the processing creating plant energy occurs. For Hemerocallis these are long parallel veined leaves. They are simple and are prototypical of the monocot leaf and especially those in the Liliaceae family. Generally the leaves are similar amongst the species with the size and relationship to the height of the scapes being the significant factors. For example H. fulva has moderately large leaves whereas H. minor has almost grass like leaves.

2.2 Inflorescence: The inflorescence is a combination of two elements, the scape and its structure and the flowering element:

2.2.1 Scape: The scape is the structure upon which the flowering structure or structures grow. The scape may be branched or unbranched.

2.2.1.1 Bracts: These are the small leaf like growths that are frequently at the dividing line between scape body and branch body.

2.2.1.2 Branch: This is the extension of a scape upon which one or several stems may extend.

2.2.1.3 Stem: This is the growth upon which the flower extends.

2.2.2 Flowering Structure: The flowering structure or structures are at the distal end of the scape and they contain the reproductive portions of the plant. They go through three phases:

2.2.2.1 Buds: The buds are the pre-bloom phase of the flower and they have certain characteristics which may help to differentiate between species.

2.2.2.2 Flowers: The flower is the most visible portion of the plant. In the genus Hemerocallis it consists of three petals, three sepals, six stamen or male pollen parts and one pistil, the female part which is divided into a sigma, style and ovary. The fertile ovary gives rise to the seeds and seed pod. The flower presents the greatest variability generally to the plant.

2.2.2.3 Pods: The pods are the fertile ovaries with seeds in the interior.

We will return to these in further detail when we look at the species but for the purpose of establishing a structure to continue this is adequate.

2.5.2 ROOTS

The roots of the different members of the genus do shown some variability. We show two of them here to demonstrate the extremes.

First is the root of H. citrina shown below? Note that these roots are long and cylindrical in shape. There is no swelling of the root and there is hair like protrusions on all of the cylindrical elements. The roots are dense and do not appear to have elements which spread out in a rhizome like fashion in a runner.

FIGURE 4 ROOTS OF H. CITRINA

The following Figure shows the roots of H. flava. Note the dramatic difference. There are bulbous sections in many of the roots which are swollen sections which are several times the diameter of the other portion of the root. It makes the root look almost like small potatoes. There are other sections which are cylindrical like H citrina but when closely investigated they are runners which shoot out and establish new plants. The bulbous roots contain a great deal of nutrients and this makes the H fulva a very hardy plant.

FIGURE 5 ROOTS OF H FULVA

2.5.3 SCAPES

The AHS defines a scape as[7]:

"The scape of a daylily is a leafless stalk which bears the flowers. Most have two or more branches, each bearing several flower buds. Below the branches, the stalks have a few leaf-like "bracts." Sometimes, a small plantlet grows at the junction of a bract and the scape. This is called a "proliferation" and can be rooted to produce another plant."

The scape is the bearer of branches, then a stem and then a set of flowers. The scape is the main body holding all the flowers. The bract, a small leaf like growth is the defining element between a scape and a branch. The branch may then break into steps, which are bractless, and the stems hold flowers.

2.5.4 BRANCHING

Branching is that characteristic of the scape which tells a great deal of the species. The bract defines the base of a branch. Some are highly branched like multiflora and some are not at all branched like dumortieri. The Figure below shows three species and their branching. The branching is the breaking off into a distinct and separate element of the scape upon which there is a set of one or more terminal flowers.

[7] See: http://www.daylilies.org/AHSfaq1.html

FIGURE 6 H. FULVA, H. AURANTIACA, AND H. HAKUNENSIS (LEFT TO RIGHT)

Another example is shown below which is H multiflora branching. This is the most branching of all the species. It is robust and at the end of each branch there are many stems with many flowers per stem.

FIGURE 7 H MULTIFLORA BRANCHING

The bract at the lower branch is shown below. There is a long slim bract seen in the middle of the branch point. These bracts are at each of the H multiflora branching elements.

FIGURE 8 H MULTIFLORA BRACT AT BRANCH POINT

The careful examination of the scape and the branch points is essential. There are many species which are non-branched. Thus the branching as identified by the displaced bract is a key identifying characteristic.

Finally we show an H coreana scape as follows:

FIGURE 9 H COREANA SCAPE

Note the bracts at mid scape and the bracts at the top of the scape. H coreana has many such bracts. Arguably the top bracts define two separate branches. As Erhardt states the "scape has numerous branches at the apex"[8]. One can see the bracts and the branches at the top of the scape branching out with the flowers on the stems.

2.5.5 BUDS

Buds are formed at the terminal ends of the scapes and it is within the bud that the flower is formed. The buds have different shapes and coloration as well as having bracts, leaf like proturbances which may be at the base of the bud.

The bud below is from H. dumortieri. It has several important distinguishing characteristics. First, the buds are sessile, or they are at the end of the scape with no branching and they have no stems or pedicels upon which they are growing. They are just there. Second, there are bracts, small leaf like growths at the base of the buds. Recall that bracts are what define the branch from the scape. Here the bract is right up against the bud. There is no stem and there is no branch. Third, there are a few buds, in the case of the Figure, four. The buds may also be slightly colored and this is slightly evident in the Figure. The coloration is seen at the distal end of the buds.

FIGURE 10 H. DUMORTIERI

[8] See Erhardt p. 39.

2.5.6 FLOWERS

The flower of the Genus Hemerocallis is a simple structure. It has three petals, three sepals, six stamens and one pistil. It is one of the simplest structures of all the flowers. The structure is consistent across all members of this genus, with the exception of the double variants of H fulva the Kwanso and Flore Pleno variants which are doubles. These are truly genetic variants and are for the most part sterile.

The picture below is for H. aurantiaca. Let us look at the characteristics of this flower:

First, color; the flower of H. aurantiaca is reddish in tone with a white mid rib, a rib down the middle of the petal and the sepal. The throat is gold toned. The H aurantiaca is an interesting plant in that it had a much more reddish flower than H fulva.

FIGURE 11 H. AURANTIACA

In the Figure below we show H. middendorfii. In many ways it is like H. dumortieri. It blooms early, it is somewhat sessile, it has bracts, but it is a chrome yellow color and lacks the brownish tines on the sepals.

FIGURE 12 H. MIDDENDORFII

The Figure below shows H. hakunensis, a branched Hemerocallis with many buds per branch. The flower is a chrome yellow.

FIGURE 13 H. HAKUNENSIS

The following Figure depicts some of the detail of the flower, namely the pistil and stamen. The ends of the stamen, called anthers are covered with pollen, the male cells. The pistil distal end is covered with a sticky fluid which will allow the pollen to adhere and then makes its way down the pistil tube to the ovary to bond with the female cell.

2.5.7 STAMENS AND POLLEN

The stamens and pollen have been used by others to identify genus other than Hemerocallis. Generally, however, in Hemerocallis, the stamen and pollen provide little

if any distinguishing a set of characteristics useful for a differentiation. We show a typical example in the picture below.

FIGURE 14 PISTIL AND STAMEN

2.5.8 PODS

The pods are the matured fertilized flower ovaries which contain the seeds.

The Figure below shows the H. altissima pod. Note that it is elongated and a light green in color and have a slightly discolored distal end. If we were to quantify the characteristics, namely provide quantified characters, we could choose the minimum and maximum diameters of the pod.

We define these as:

$d_{\min}, d_{\max}$ Respectively. Now these measurements are themselves in a distribution within a species. This for H. altissima we may have:

$$d_{\min} = m_{\min} + \sigma\Delta$$

Where:

$$m_{\min} = \sum_{k=1}^{N} \frac{d_{\min}^{k}}{N}$$

and

$$\sigma^2 = \frac{1}{N-1} \sum_{k=1}^{N} \left(d^{k}{}_{\min} - m_{\min} \right)^2$$

And Δ is a zero mean and unit variance Gaussian random variable. That is the diameter is a Gaussian with a mean and a standard deviation. We have performed many tests and

this seems to be confirmed in our analysis. In fact one can do some classification on pods using the pod measurements.

FIGURE 15 H ALTISSIMA POD

Now look at the H. citrina pod. This is shown below. The distal end has the same coloration but the pod is fatter towards the distal end and the ratio of $\frac{d_{\max}}{d_{\min}}$ is not well defined due to this no ellipsoidal shape.

FIGURE 16 H. CITRINA POD

The same shape variation is seen for the H. thunbergii pod shown below. Note that the pod here is characterized by a more complicated set of quantities. For example we may consider the following.

FIGURE 17 H. THUNBERGII POD

Consider the geometry of the pods shown below. If we know we have an ellipsoid then we need only minimum and maximum diameters. If, however, we have a shape more like the second form we need to know the minimum, the maximum and the offset of the maximum. Thus we need three parameters. If the pod is ellipsoid, we know a priori that the offset is equal to one half the maximum.

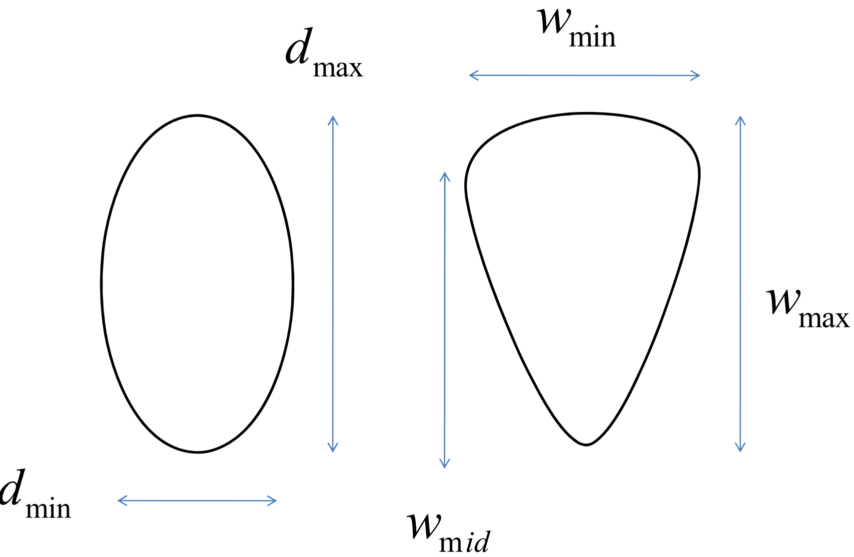


FIGURE 18 GEOMETRY OF PODS

Now it is possible to perform a classification using the pod shape. We shall discuss classifications latter in the book but this is a good point to demonstrate that there is a commonality within the species and a variation between the species by looking at pods alone. The above examples of pods clearly show that the trained eye can discriminate between pods. The H citrina pod, for example, has the dark tip which is another distinguishing characteristic.

2.5.9 Seeds

The seeds are the final evolution of the flower from bud, to flower, to pod to seed. Admittedly we are looking at different parts in each step but from a character or characteristics perspective we are inspecting another element. The following shows some variations that one may see in the seeds.

In the Figure below we have presented six species seeds for comparison. H. coreana and H. citrina have the largest seeds, 5-7 mm and H. minor the smallest at 2-3 mm. Some are round, some oval, and one could then also characterize the seeds by shape. Such measures as:

Minimum Diameter: The smallest diameter measurable across the center.

Maximum Diameter: The largest diameter measured in a similar manner.

Ratio of Max to Min Diameter: Just the ratio of the above.

One could continue with such metrics and they could then be used in a subsequent classification. However, in looking at the samples shown below one could assume that there is small intra species variability and moderate interspecies variability. That is if we assume that the average minimum diameter is:

$$m_{Species1}$$

And we assume that the distribution if Gaussian about this mean with density:

$$p(d) = \frac{1}{\sqrt{2\pi\sigma_2}} \exp(-\frac{1}{2}\left(d - m_{Species}\right)^2 / \sigma^2)$$

The following shows the seeds of several species:

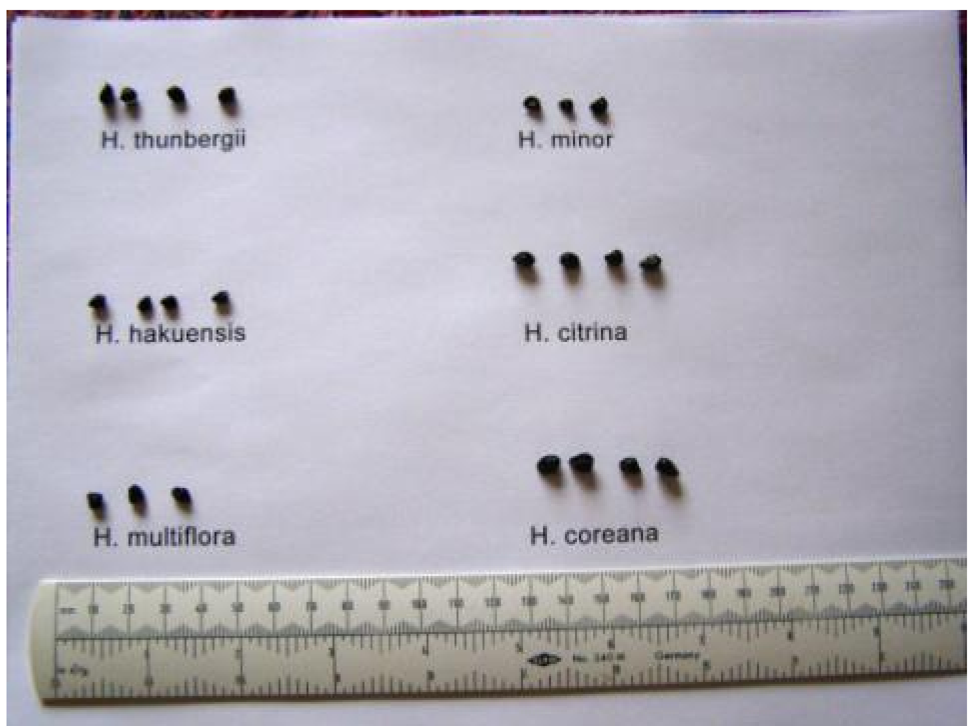


FIGURE 19 SEEDS FOR SIX SPECIES

The above depicts the seeds for six species. We have looked at the analytical study of the seeds as discriminators for species but unlike the pods the seeds are much more difficult to use for that purpose. H minor is quite small and H coreana are quite large. The shape is quite variable and difficult to maintain and closeness within a species. In

our experience the seed is a poor discriminator for species. We have gathered a few H fulva sempervirens seeds but we have no fertile H fulva seeds alone.

2.6 SPECIES

The species Hemerocallis is indigenous to Asia, specifically China, Korea, Japan, and Eastern Russia. It is a mountainous plant and is generally quite hardy and very resistant to diseases. Since the late 19th century there has been a great deal of hybridizing of the plant. Thus, for just over one hundred years, hybridizers have been cross the species and their descendents to create a wide variety of new and innovative hybrids.

From the species which is predominantly yellow, orange and a brownish red color, comes a wide variety of forms and color. Bright reds, purples, shades of gold, doubles, plants with eyzones and plants with spider like form and shape. In this section we review the genus and its associated species and then we look at some of the hybrids.

The Genus Hemerocallis has a dozen or more species.[9] The identification of the species is still somewhat in flux. One of the earliest classifications was done by Stout in the late 30's and still stands with some modifications. There are many others who have proposed alternative classifications but when one looks at the literature one seem many differences and a few commonalities. We will in this paper not focus on a definitive classification but use several of the better defined species to make the point.

Below we have shown several species and their variation. One must recognize several factors even in a species;

1. Species are geographically clustered. Thus H. citrina in one place will look like H. citrina in another but there may well be differences.

2. There has been some work on the genetic diversity within and between species. There is still a great deal more to be done.

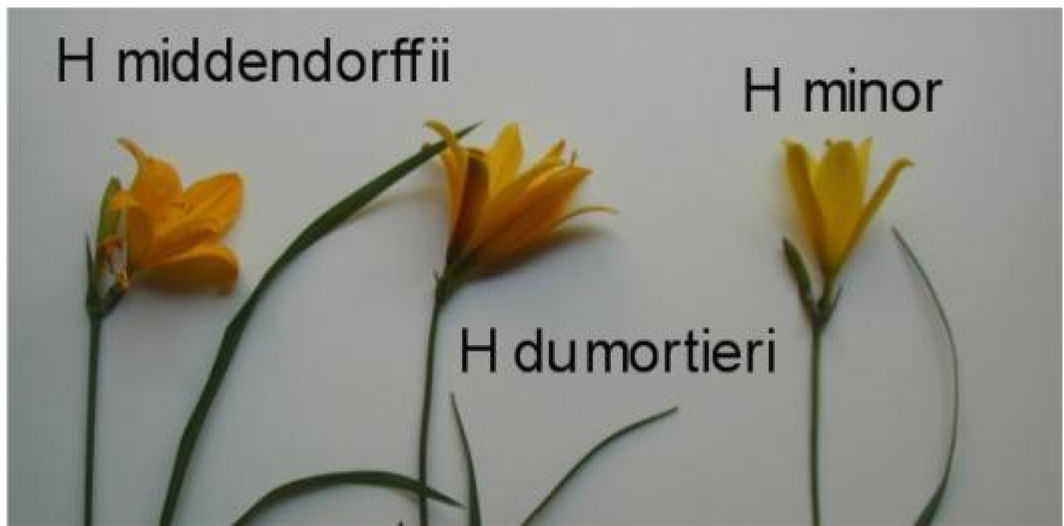
3. Many species are self sterile, such as H. citrina, but can be crossed with other species to create hybrids.

4. Some species like H. fulva Europa are triploid and are sterile and propagate via a vegetative process.

[9] See the papers by Schabell. They are an excellent historical collection of the original works characterizing the species. The work by Stout still remains per-eminent. The work by Erhardt is somewhat useful but I have found inconsistencies and in addition it is extremely difficult to see an overall structure.

5. Variability exists within species and within the same geographic area, and one sample of a species may not look exactly like another from the same location, however the variation is a micro variation, one could still identify the species from the collection of phenotypic characteristics. None of the species expresses the characteristic we see in many of the newer hybrids, and that will be a question key to this analysis.

Note when looking at the species flower colors we see yellow, reds, some darker brownish reds, and orange. There is some variation of color within a species.[10] The three plants below show three of the earliest blooming plants; H middendorfii, H dumortieri and H minor. They are all sessile, namely have no discernable branching and further the stem is also non-discernable, and there are bracts at the base of the flower. H minor has a grass like foliage and the scape tends to flop down on the ground. H dumortieri has brownish sepals and yellow petals and H middendorfii is an erect scape with strong yellow flowers. H middendorfii is generally the first plant to bloom.



The above also shows the branching habit of the flower. The three shown are all early blooms. The H. coreana species is shown below. The two plants have been obtained from two sources. Note the difference in color. The question is, are these two plants of the same species, have they been hybridized, or is one a species and the other a hybrid. In fact one may even ask are the related at all.

We will now look through the twelve species which we have been able to collect and grow successfully. The first two are H altissima and H aurantiaca. We show the flowers below. H altissima is a very late blooming plant, generally the latest except for H fulva

[10] The recent paper by Tompkins is useful since it uses the AFLP approach to determine a broad base of cross species variability as well as geographical variability.

sempervirens, which we shall discuss latter. It has some variability in the plants we have received.

Some are quite tall and others are of medium height. H aurantiaca has been listed as a separate species from H fulva and in the plants we have grown they show a strong reddish tint separating them from the orange red of H fulva. We will provide additional differences in the Appendix of this Chapter. However, one always looks to these plants and must ask if there are different species or just variations. We will answer that question latter after having addressed the gene issues. Note also in H aurantiaca the presence of a mid rib on both sepals and petals.

FIGURE 20 H ALTISSIMA

FIGURE 21 H AURANTIACA

The next two plants are H flava and H citrina. H citrina is a night blooming flower and is highly branched and very fragrant. We grow H citrina from six different sources in east Asia. They tend to bloom at differing times, some in mid July and some two weeks latter. They have varying heights and they all seem to bloom at night. We have some slight success in seed setting but it is not as strong as other species. H flava is now called H lilioasphodelus and it is recorded as being one of the first species brought to Europe. It is a clear yellow flower of medium size. It also has fragrance and it blooms in the day time.

FIGURE 22 H. FLAVA

FIGURE 23 H CITRINA

H coreana is a plant which we find to be an aggressive grower in clumps and the flowers are almost sessile. We have several variants and they tend to have some difference in the length of the bracted stems. There is a mild difference in color from a yellow to a gold. Again one must ask if these may be hybrids from the wild or is there an

intraspecies variation. Frankly the question is not readily answered at this time. The h dumortieri is an early blooming plant. I is a slight bi-color having yellow petals and sepals with bronish tinge on the outer side.

FIGURE 24 H COREANA

FIGURE 25 H DUMORTIERI

The picture below shows the H coreana branching. It appears to be no real branches and just short stems all with bracts. These are two separate H coreana plants from various sources. We find that H coreana has tremendous seed setting capability and we propagate them aggressively. They are an excellent horticultural plant and we have also used them in hybridizing in an attempt to introduce their strong growing characteristics.

FIGURE 26 H COREANA BRANCHING 1

FIGURE 27 H COREANA BRANCHING 2

The H. coreana color variation seems to be substantial across many of the plants we have seen in the United States. The left one above has been growing for several years, Right one is recent acquisition. Both have bracts, large ones which show it to be most likely H. coreana. Again we ask: Why the color difference? Is it a variety, geographically different part of species, early color to change latter?

H. coreana is also an evergreen plant. Here we have even new growth after a fairly cold winter. This seems to be the first species to start leaf growth in late February.

The next two are H fulva and H hakunensis. H fulva has a few varieties which we have grown including the double Kwanso and Flore Pleno and the late bloomer H fulva sempervirens, and a variegated leaf double, which has a green and white striped leaf. Some of the H fulva varieties are fertile, the common one seen is generally not and it is the H fulva Europa. H hakunensis is a branched yellow flower which blooms mid season and is a strong bloomer. It sets seed but nowhere as aggressively as H coreana. H fulva has am orange brown color to make it different from H aurantiaca which is clearly reddish brown.

FIGURE 28 H FULVA

FIGURE 29 H HAKUNENSIS

The next two are H minor and H middendorfii. H minor is almost grass like; it also re-blooms in the fall. Thus the re-bloom characteristic is quite evident in H minor. It sets seed and it has leaves that look like floppy long grass, and its scape also tends to be long and floppy. H middendorfii is the earliest bloomer. In region 6-5 where we grow it, if it is in the sun all day it will bloom in early April. It blooms two weeks earlier than H minor. H middendorfii is a strong scaped plant, but the flowers are sessile and un-branched.

FIGURE 30 H MINOR

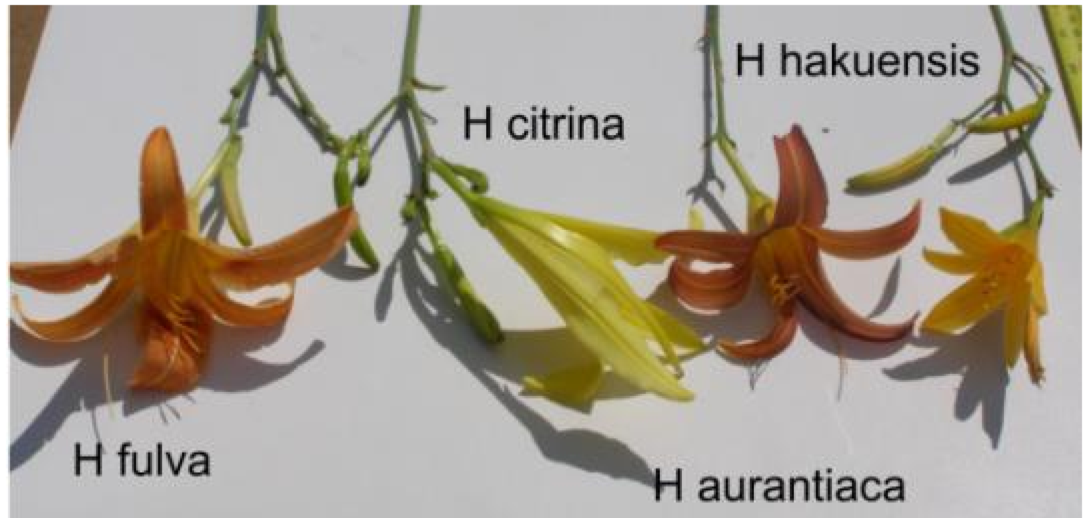
FIGURE 31 H MIDDENDORFII

H multiflora is a highly branched flower with yellow petals and sepals. We have grown a three different source plants, two from Dr. Apps, and they show some variation in color from yellow to an orange tint. Again we do not know if some of these may have been hybridized in the wild or if this is a natural variation. One can see the difference in the plants when they adjoin one another. H thunbergerii is also yellow and less branched than H multiflora. Both are yellow.

FIGURE 32 H MULTIFLORA

FIGURE 33 H THUNBERGERII

We can compare some of these species as shown below. The following picture is of four flowers from citrina, aurantiaca, fulva (species NOT Europa), and hakunensis.



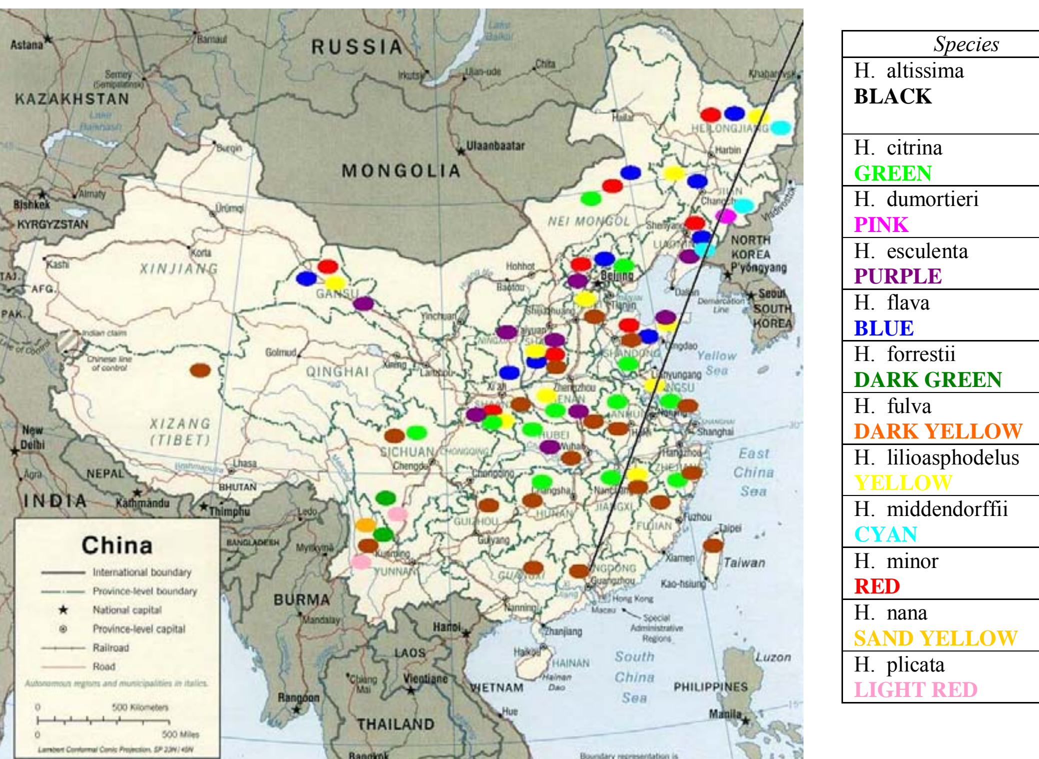
Note the branching. H aurantiaca is clearly reddish in tint and H citrina is closing from the night before. H hakunensis is branched.

We now present a graphical description of where the various species may have originated from. The geographic distribution of the species in their original locations in China is shown below.[11] Note the many locations and that most of them are mountainous. There are other locations in Russia, Korea and Japan. Generally they are found in mountainous regions.[12]

[11] This has been prepared by Yongang Wen, of MIT, one of my graduate students.

[12] We have found that our best results are on the lower slopes of the White Mountains in New Hampshire. The climate is similar to the areas where the species come from originally. It is cool, and in Zone 4b with lowest temperatures at -35 F and is sand soil with 52” to 58” of rain per year.

Origin



To address the entirety of the species we must look at all the putative species, several of which we cannot grow in our region due to the extreme cold. We see temperatures as low as 0F in the winters. The current list of all Hemerocallis species is as follows:

Species	*Comment*
1. H altissima	This is a well established species available generally.
2. H aurantiaca	This is a well established species available generally.
3. H citrina	This is a well established species available generally.
4. H coreana	This is a well established species available generally.
5. H darrowinina	
6. H dumortieri	This is a well established species available generally.
7. H esculenta	
8. H exaltata	
9. H forestii	
10. H fulva	This is a well established species available generally.
11. H graminea	This is apparently not cold tolerant.

12. H hakunensis	This is a well established species available generally.
13. H honngdoensis	
14. H lilioasphodelus	This is a well established species available generally. Also called H flava
15. H littorea	
16. H longituba	
17. H micrantha	
18. H middendorffii	This is a well established species available generally.
19. H minor	This is a well established species available generally.
20. H multiflora	This is a well established species available generally.
21. H nana	This will not grow in Regions 8 and lower. It is not cold tolerant.
22. H pedicellata	
23. H plicata	
24. H taeanensis	
25. H thunbergii	This is a well established species available generally.
26. H yezoensis	We tried to grow this species here but it did not survive.

There is a total of 26 named species of which we currently grow twelve. Two of the 14 we do not grow will not survive this far north. The others seem to be located in isolated regions.

We have also created a data base for the phenotypic characteristics of the plants in our current collection. This is shown below. One could use this data to create a clade analysis and then also create a key. We have done that in another report and we feel that the result should be considered as preliminary.

If we were to take the various authors who have attempted to characterize the species we obtain the chart shown below. One thing evident in the chart is the lack of agreement. Again we believe that agreement can only be obtained after a detailed genetic analysis.

Name	*Stout (1934)*	*Erhardt (1992)*	*Plodeck (2003)*	*Munson*	*Hortus Third (1976)*	*Peat & Petit (2004)*	*Grenfell (1998)*	*Petit & Peat (2000)*	*PFAF (2000)*	*McGarty*
H altissima		X	X	X	X	X	X	X	X	X
H aurantiaca	X	X	X	X	X	X	X	X	X	X
H citrina	X	X	X	X	X	X	X	X	X	X
H coreana		X	X	X		X	X	X	X	X
H darrowinina						X	X		X	
H dumortieri	X	X	X No Picture	X	X also calls it H sieboldii	X	X	X	X	X
H esculenta		X	also he calls dumortierii and middendorffii v esculenta in this species	X		X	X	X		
H exaltata	X	X	X	X			X	X	X	
H forestii	X	X	X No Picture	X	X	X	X	X	X	
H fulva	X	X	X		X	X	X	X	X	X
H graminea				X	X also H dumortieri and H minor		X	X	X	
H hakunensis		X	X	X		X		X	X	X
H honngdoensis						X				
H lilioasphodelus	X H flava	X	Also H flava	X Calls it H flava	X	X	X	X uses H flava	X	X
H littorea				X				X	X	
H longituba				X						
H micrantha		X	X no picture	X		X			X	
H middendorffii	X	X	Also H dumortierii v middendorffii	X	X	X	X	X	X	X
H minor	X	X	X	X	X	X	X	X	X	X
H multiflora	X	X	X	X	X	X	X	X	X	X
H nana	X	X	X No Picture	X	X	X	X	X		
H pedicellata		X	X No Picture	X		X			X	
H plicata	X	X	X No picture	X	X	X	X	X	X	
H taeanensis						X				

H thunbergii	X	X	X also H serotina and H sulphurea	X	X but calls it H serotina	X	X	X	X	X
H yezoensis		X	X also H flava v yezoensis			X	X		X	

Some authors have placed these species in groups. We have shown this in the following Table. Erhardt seems to be setting the standard but there are several inconsistencies in his approach, Peat and Petit appear to be repeating Erhardt. There is no true well established and accepted classification, however.

Name	Stout	Erhardt	Plodeck	Munson	Hortus Third	Peat & Petit	Grenfell	Petit & Peat	PFAF	McGarty
Groups										
	1. Forked; Scapes < Leaves (nana, plicata, forestii)	Fulva (aurantiaca, fulva)				Fulva (aurantiaca, fulva, hondoensis, taeanensi)				
	1. Forked; Scapes > Leaves (flava, minor, thunbergerii, citrina, fulva, aurantiaca, exaltata, multiflora)	Citrina (altissima, citrina, coreana, lilioasphodelus, minor, pedicellata, thurbergii, yezoensis)				Citrina (altissima, citrina, coreana, lilioasphodelus, minor, pedicellata, thurbergii, yezoensis)				
	2. Unforked, Unbranched (dumortierii, middendorfii)	Middendorffii (dumortieri, esculenta, exaltata, hakunensis, middendorffii				Middendorffii (dumortieri, esculenta, hakunensis, middendorffii)				
		Nana (forestii, nana)				Nana (darrowinia, forestii, nana)				
		Multiflora (micrantha, multiflora, plicata)				Multiflora (micrantha, multiflora, plicata)				

2.6.1 Buds

The buds are the parts of the plant that we see just before the flower is produces. The buds have specific characteristics which allow for characterization and classification. The following is a summary of several of the species.

Consider the buds shown below. The first is H altissima from Olallie. It is shorter than some alleged altissima plants and blooms much too early to be a true altissima, about a month earlier than what is generally accepted. It is suspected that it is not the more classic altissima as the others altissima that are 70-80" in height and bloom later.

The second is H multiflora. There is great branching and a great many buds before the flower emerges. The branches mature well before flowering. This is classic multiflora. Note the aggressive branching. H multiflora is frequently used in hybridizing to achieve a highly branched hybrid.

FIGURE 34 H ALTISSIMA

FIGURE 35 H MULTIFLORA

We can now look at additional buds. The first is an H. minor obtained from Apps and attributed to be from Siberia. This is a late bloomer for a minor and the leaves seem a bit too erect as does the inflorescence. The second is a classic H flava form from Apps. Note the bracts and the almost sessile ends.

FIGURE 36 H MINOR

FIGURE 37 H FLAVA

We now look at H citrina and H hakunensis. This is a classic H citrina. There is slight branching and 3 buds per branch. The ends of the citrina buds have a dark brown or red color. The H hakunensis is splayed at the top also with three buds per set.

FIGURE 38 H CITRINA

FIGURE 39 H HAKUNENSIS

The following two are for H fulva and H aurantiaca. H fulva buds show the orange color even through the bud whereas the H aurantiaca can see the reddish color.

FIGURE 40 H FULVA

FIGURE 41 H AURANTIACA

2.6.2 PODS

Pods are also descriptive of a species as we have discussed. We look here at eight species and discuss them in some detail. Again we detail these again in the Appendix at the end of the Chapter

FIGURE 42 H CITRINA

FIGURE 43 H MINOR (APPS)

The next two are altissima and aurantiaca. They are both long and thin and they are differentiated by the fact that altissima has a brownish sport at the terminal end. The H aurantiaca pods are dark green and remain that way for a long period of time.

FIGURE 44 H ALTISSIMA (OLALLIE)

FIGURE 45 H AURANTIACA

H coreana is a strong and aggressive seed grower and the pods are large and round with the distal tip being much larger that the proximal side. They frequently are quite extensive since they appear to self pollinate greatly in this area. The hakunensis pod is larger across its center; it is ovoid in shape and is generally a solid green mass until it finally dries out.

FIGURE 46 H COREANA

FIGURE 47 H HAKUNENSIS

The H multiflora pod is also ovoid and it appears to be a good seed setter but not as good as H coreana. In contrast the H thunbergerii is more like H coreana having an enlarged distal end to the pod.

FIGURE 48 H MULTIFLORA

FIGURE 49 H THUNBERGERII

2.6.3 Scapes

The following are some detailed descriptions of H coreana. The picture below is H coreana. Note the bracts just above mid scape. There also is no branching in the flower and the color is a bright yellow, almost chrome. This picture with the shadow gives a good presentation of the H coreana.

FIGURE 50 H COREANA

FIGURE 51 H COREANA BRANCHING AND BRACTS

The full scape for H coreana is shown above It shows the bracts and the budding and the splitting of the buds at the end of the scape.

2.6.4 Seeds

The seeds as we have noted earlier have less individual identity. For the species we have detailed some of them here.

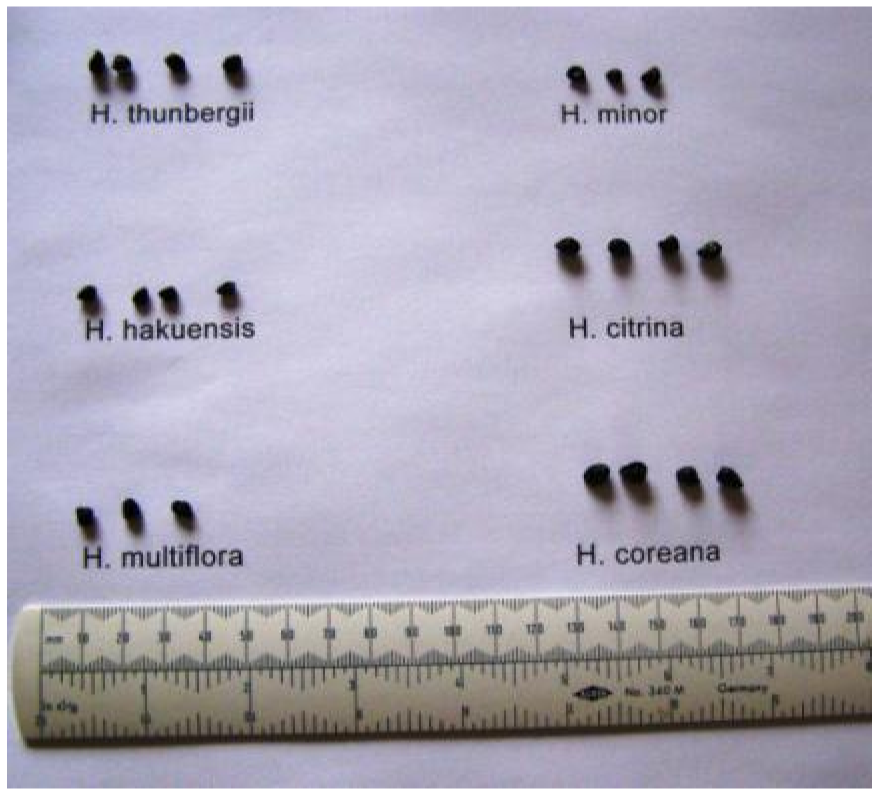


FIGURE 52 COMPARISON OF SEEDS

Let us look more closely at H. citrina. This is shown below.

FIGURE 53 H. CITRINA SEED

Note that the seed has various shapes. One is much more round and the others have some indentation. These seeds have come from a cold stratification so there may be some collapsing from the dehydration which may occur. We will show fresh seeds as well. Below we shown H. coreana in a close up.

FIGURE 54 H. COREANA SEED.

This is a round seed with a maximum to minimum ratio of between 1.5 and 2.0 and there are seeds showing some collapse due to dehydration. This

The next Figure is for H. thunbergii.

FIGURE 55 H. THUNBERGII SEED

This seed, H. thunbergii, shows a similar shape, but it is not as elongated, on average.

Thus the characteristics of seeds which we may use are their minim and maximum diameters, and their ratios. Other

2.7 Early Bloomers

As a prelude to establishing a key for classifying the species we look at the three early bloomers. In our view a key will be based first on the most obvious characteristic and then on the lesser obvious. For us the most obvious is the time of bloom.

First we look at the H dumortieri species as shown below:

Note the following:

1. On the pictures bellow one can see that the back of the sepals we see the dark reddish brown tint. The petals are pure yellow.

2. The buds are clusters with bracts at the base.

3. The H dumortieri seems to bloom a week after the H middendorfii. The H middendorfii however is a great grower from seeds and flowers now in large clumps.

Now we have also tried to compare the flowers. The three flowers are shown to the left and also below. From the left we have H middendorfii, H dumortieri, and H minor. Note first the color. The middendorfii are an orange color. This color seems to breed true on the plants which I have it is a good plant to grow from seed, in fact the best I have seen. I now line the driveway with them and they have a large group of flowers on each plant. Then the dumortieri which is yellow, and slightly darker yellow than the minor. The minor is a pale yellow with no tints on either sepals or petals. Also minor has grass like leaves.

Here we can see minor on left, then dumortieri, and then middendorfii. The tint on the sepals of the dumortieri is quite evident in this picture. Also we see the middendorfii petals and sepals opening in a wide bursting manner on many of the plants. There is also another view of the above but with the scale along the horizontal.

H dumortieri has bloomed. It is shown to the left. Note the dark reddish brown sepals and the chrome petals. The difference between dumortieri and middendorfii should be clear with this one characteristic.

The dumortieri seen above also shows several characteristics. First there are bracts mid scape as shown on bottom right. Second, the buds are sessile, no branching, and the

buds show the reddish tint as well. It is very prominent just before flowering. H minor is generally the second species to bloom.

Also we see that H minor is really a night bloomer and is quite fragrant. The plant is a very sensitive bloomer, just a few scapes per plant and the plant does not seem to propagate as well as others. We can also see that there are just two buds per scape, with a small bract and the buds are fairly sessile. I have noticed that it blooms at the end of the day and the peak openness is just after sunset. I have been hybridizing these as well.

The minor has a very slight branching and is not clearly sessile. The flower opens greatly at night and closes down in the day time. We notice only two buds per scape and the strong yellow is very attractive.

H middendorfii blooms from early April thru mid May. The flower is a chrome yellow as shown below and is quite sessile.

The flowers are shown above and to the left. The chrome yellow is quite prominent. The flowers are sessile; all clumped to the end of the scape, there is no branching. There are small bracts at the base of the buds and three to four flowers per scape. These plants have all been grown from seed. The H middendorfii seems to grow the best from seed of all the species. We have been able to grow hundreds of these plants and we are now testing them in New Hampshire. The New Hampshire plants are about a week to ten

days behind these. However the other New Hampshire plants are three to four weeks behind New Jersey. This may imply that the Hemerocallis is more light dependent and not as reliant on heat as many other species. An interesting observation.

H minor blooms about a week later than H middendorfii. The flower has a clear yellow like color, there is no clear branching and the leaves are quite narrow and grass like. This is the first species to bloom. We have transplanted a large collection of minor to a naturalizing and species display garden and they do not seem to bloom. H minor seems to be a bit fragile in its ability to move. We have planted these in our New Hampshire gardens and will see how well they do in Zone 4.

Buds from the species plants are shown below. It is useful to notice the difference between species at this early stage. There are four species in bud; minor, dumortieri, middendorfii, and minor.

2.8 Later Blooms

The later bloomers come about a month after the early bloomers. This means the remaining nine species. The H flava from Apps blooms in mid-summer. It is shown below.

The above shows the H flava that we received from Apps. This is the first flava plant and it does appear to look a great deal like the minor but the difference is that the leaves are much wider and the scape stands upright. Since this is the first year of a bloom the timing may not be normal but if it is then the timing is about two weeks later than minor.

The above shows the buds for H flava. There are small stems extending above the scape and they appear in groups of 2. Also there is a bract below the budding area.

H minor can be compared as we do below.

FIGURE 56 H MINOR BUD

FIGURE 57H FULVA BUD

The above shows the bud for the H minor from Apps which is a Siberian H minor. This is about three weeks behind the H minors I have in the garden from years past. It may take another week to bloom. Like all the other recent species this may not be blooming at a normal time because of the transplant shock.

The above is the bud for H fulva, the common daylily. Note the large bract that is common to this sterile plant. They are all in bud and we use them for naturalizing, although they are now considered an invasive species.

We have two species received from Apps last year. They are H flava and H minor (Siberia).

FIGURE 58 H FLAVA (APPS)

FIGURE 59 H MINOR (SIBERIA APPS)

The H flava shown above has a brownish color on the sepals. We had not seen this on the first bloom. It is also quite fragrant and blooms initially in later afternoon. The flower lasts more than a day and is a strong yellow color. The flowers are sessile but there is the formation of a secondary sessile bud below the first. They appear in groups of four. The H minor from Siberia is also shown above. Note the number of ants. It must be secreting sugars to attract the ants. This was taken at about 19:00. Also note the bracts on the buds. To the left is a second set of buds as well.

2.9 Keys and Classification

Keys have been used to help identify the different species. The key is in many cases different from a classification. A classification, which we shall detail later, considers the evolutionary or genetic relationship between species and groups them in a manner consistent with their evolutionary proximity. The classic Cronquist systematics classification is a classic example. There are many others. The key helps the individual identify a field plant by its characteristics. A key does not have to relate to a classification, and in fact they may be in contradistinction to one another. In this section we take all the species including those we have not discussed and provide several keys for consideration.

2.9.1 Simple Key to Twelve Species

The first key is one we have developed based upon our own experience. Keys to be useful must assist the user to identify based upon the most obvious characteristics first and then drill down in detail. In our case we look at the early bloomers and then the mid and late bloomers. The time of bloom may very well also correlate to the genetic proximity since it is unlikely that H middendorfii would cross with H fulva sempervirens

since they bloom four months apart. Pollen just does nothing around on the bee that long.

Thus the following two keys demonstrate the early bloomers and the later bloomers. That is the first bifurcation.

Having laid out the early bloomers we can then address the later bloomers as shown below. This builds on the detail we presented in this Chapter.

2.9.2 Stout Key

The earliest key was developed and published by Stout and is shown below. His major differentiator was sessile versus non sessile flowers. The two Figures below show this key. Note that he calls H minor a branched species and we see it as minimally branched. We also find H minor and H flava to bloom at substantially different times than Stout. We grow the plants at about thirty miles difference from each other albeit eighty years later. Thus we suspect micro climate differences but not major bloom time differences.

For the two which Stout calls unbranched or sessile, he differentiates H dumortierii as having a brown tinged bud. The better identifier in our opinion is the one stressing the sepal brown on the outside, it is an immediate identifier.

2.9.3 Keys to New Species

There have been several new species identified and we summarize them here. We have had no experience with any of them. They are. H. taeanensis and H. hongdoensis and the details on these are shown below.

Character	H. hakuunensis	H. thunbergii	H. middendorffii	H. hongdoensis	H. taeanensis
Shape of inflorescence	mainly 2–4 dichotomously branched (rarely Y-type)	mainly 2–4 dichotomously branched (rarely Y-type)	capitate type	Y-type (trichotomous)	Y-type (rarely branched twice)
Odor	absent	strong	faint	absent	absent
Shape of roots	some slightly enlarged	slender, cylindrical	slender, cylindrical	highly enlarged	some slightly enlarged
Flowering time	diurnal	nocturnal	diurnal (extended flowering)	diurnal	diurnal
Flowering period	July to August	middle July to middle August	middle May to June	last July to August	middle May to June
Bracts below inflorescence	present (1–2)	present (1–2)	absent	present (1–2)	absent (rarely 1)

The following is a proposed key for identifying these new species amongst the existing ones. This has been proposed based upon Korean species. Although we grow them at the same latitude and with generally the same growing conditions as one would find in Korea, we do not see H minor as an evergreen. In fact the only evergreen we have is H coreana. In addition H nana will not survive in our northern regions. The two new species seem close to H hakunensis having the same color but only scape and branching differences. One continues to wonder why these should be new species rather than variants but not having them limits the effort.

1. Leaves evergreen; flowers light-yellow . . . . *H. minor*
1. Leaves die in winter; flowers orange-yellow or golden-orange.
    2. Flowers golden-orange; perianth with short tube, sometime not evident.
        3. Flowers 3 or more; scapes 5 X longer than flowers.
            4. Leaves 10-21 mm long, plane; perianth tube 1 cm long. ………….. *H. forrestii*
            4. Leaves 3-9 mm long, folded; perianth tube 1.5-2 cm long. .. ……… *H. plicata*
        3. Flowers 1 or 2; scapes short, 2 4 X longer than flowers . . . . *H. nana*
    2. Flowers orange-yellow; perianth with evident tube 1.0-3.5 cm long.
        5. Plant small, 35-70 cm tall; leaves 0.5-1.4 cm wide; scapes slender, 2 mm wide at base. . . *H. taeanensis*
        5. Plant robust, 32-150 cm tall; leaves 1.0-3.0 cm wide; Scape robust, 5-10 mm wide at base.
            6. Scape with a dichotomously branched inflorescence; flowers large, 10-15 cm long; roots highly inflated. .. . *H. hongdoensis*
            6. Scape with 2-3 dichotomously branched inflorescence; flowers relatively small, 6-12 cm long; roots slightly inflated. *H. hakunensis*

2.9.4 Erhardt Key

The next attempt is by Erhardt. Erhardt is blatantly critical of Stout and in his abrupt Germanic fashion dismisses Stout out of hand. Erhardt also propose more of a classification rather than a key. Unfortunately a classification requires some basis akin to Cronquist et al but here Erhardt dismisses this either out of ignorance or arrogance, it appears to this author. In addition there are some apparent internal inconsistencies in his work. We present the details below.

Group	Species	Characteristics
Citrina Group	H. altissima H. citrina H. coreana H. lilioasphodelus (H. flava) H. minor H. pedicellata H. thunbergii H. yezoensis	Scapes Branched Blooms in evening Fragrant Mostly Yellow Long Perianth
Fulva Group	H. aurantiaca H. fulva	Roots have spindle shaped swellings Blooms are brownish-red
Middendorfii Group	H. dumortieri H. esculenta H. exaltata H. hakuunensis H. middendorffii	Scapes are not branched Bracts short and broad and do not overlap Blooms are orange
Multiflora Group	H. micrantha H. multiflora H. plicata	Scapes are many branched Flowers on short stalks and small with tubes
Nana Group	H. forrestii H. nana	Scapes less than 20 in Perianth tube short Veins on perianth not branched Plant not winter hardy

2.9.5 Plodeck Key

Plodeck has performed a brilliant task of collecting information on all the species and we defer to his work for the detailed references, current and historical, relating to the species. He has not finalized his key but his summary is below. He appears to follow Erhardt but is much more amenable to Stout and others before him.

Group	Species	Characteristics
Citrina Group	H. altissima H. citrina H. coreana H. lilioasphodelus (H. flava) H. minor H. pedicellata H. thunbergii H. yezoensis	Flower color: mostly yellow Flowering habits: nocturnal Branching: branched Roots: others: flowers are fragrant with long perianth tubes
Fulva Group	H. aurantiaca H. fulva	Flower color: brownish-red (fulvous dye) Flowering habits: diurnal Branching: branched Roots: spindle-shaped swellings others:
Middendorfii Group	H. dumortieri H. esculenta H. exaltata H. hakuunensis H. middendorffii	Flower color: orange Flowering habits: diurnal Branching: not branched Roots: others: bracts are mainly short and broad, but overlap
Multiflora Group	H. micrantha H. multiflora H. plicata	Flower color: orange, orange-yellow Flowering habits: diurnal Branching: many branches Roots: others: flowers on short stalks, smaller than 7 cm, tubes less than 2 cm long
Nana Group	H. forrestii H. nana	Flower color: orange, orange-yellow Flowering habits: diurnal Branching: many branches Roots: others: flowers on short stalks, smaller than 7 cm, tubes less than 2 cm long

2.9.6 Matsuoka and Hotta (1966) Lists Species

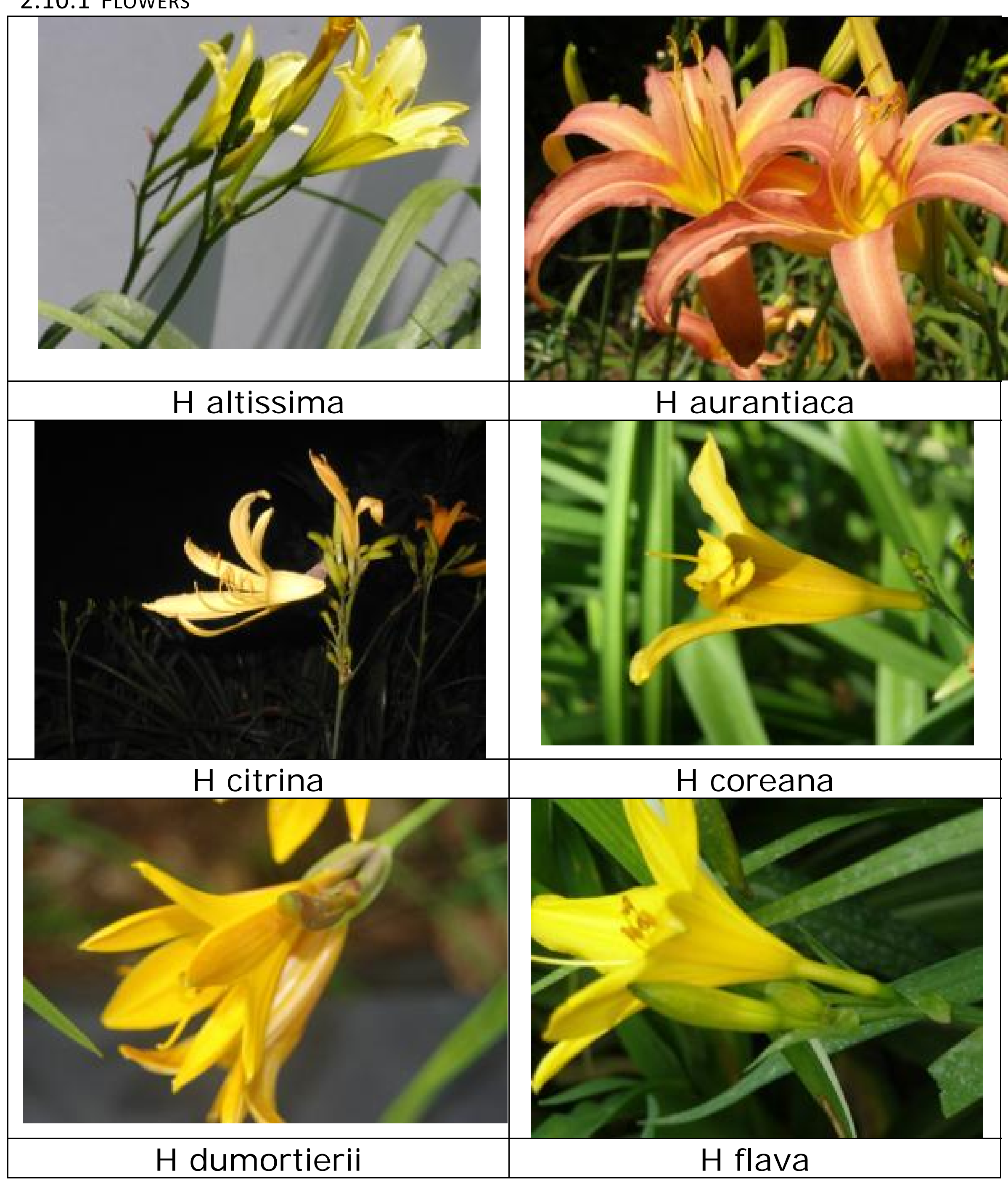
Another listing of species is due to Matsuoka and Hotta in 1966. We list them below. Unlike the list of twenty six we presented earlier this contains thirty. The major difference is the inclusion of fulva varieties as species. They list seven fulva varieties as separate species and thus are we combine then into one species we have twenty four. Then, since we added the two new ones and we are at twenty four species.

Correct Name	Number
H. altissima	1
H. aurantiaca	2
H. citrina	3
H. coreana	4
H. darowiana	5
H. dumortieri	6
H. esculenta	7
H. exaltata	8
H. forrestii	9
H. fulva	10
H. fulva Europa	11
H. fulva 'Flore Pleno'	12
H. fulva 'Kwanzo'	13
H. fulva 'Kwanzo Variegata'	14
H. fulva rosea	15
H. fulva var. sempervirens	16
H. graminea	17
H. hakuunensis	18
H. hongdoensis	19
H. lilioasphodelus	20
H. micrantha	21
H. middendorffii	22
H. minor	23
H. multiflora	24
H. nana	25
H. pedicellata	26
H. plicata	27
H. taeanensis	28
H. thunbergii	29
H. yezoensis	30

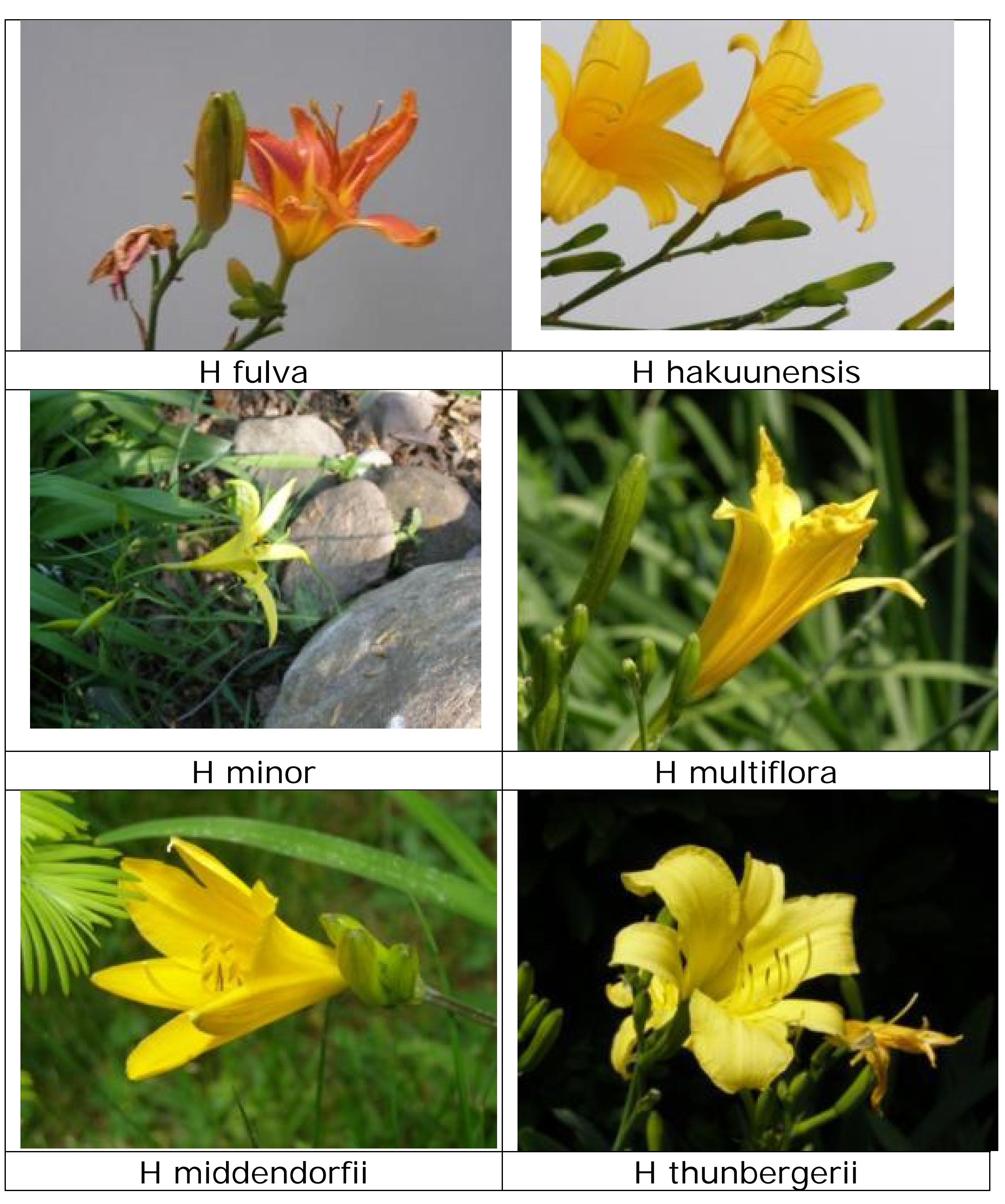
2.10 Appendix: Summary of Species

The following Tables depict the detailed summary of the twelve species we have discussed herein.

2.10.1 Flowers

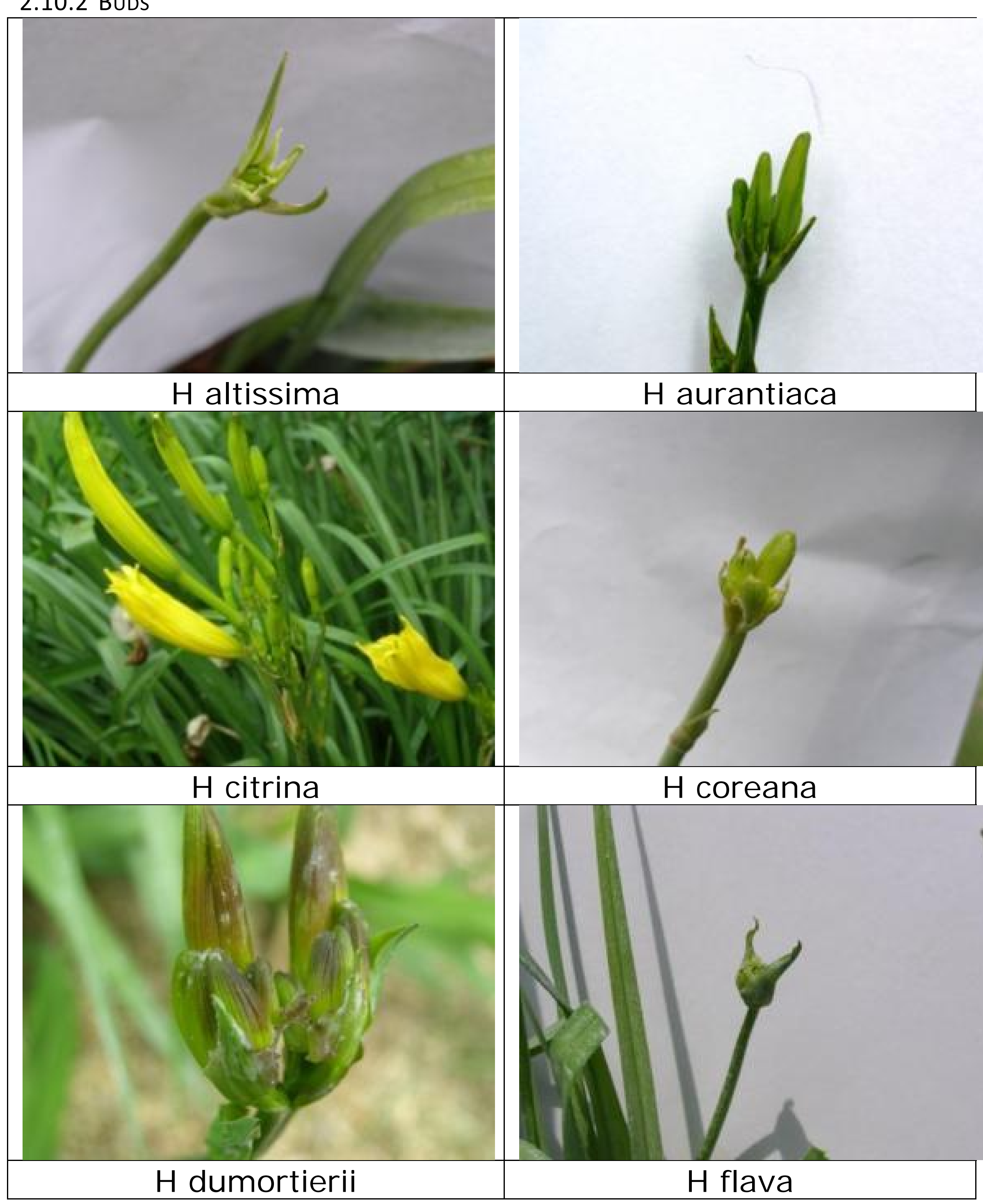


H altissima	H aurantiaca
H citrina	H coreana
H dumortierii	H flava

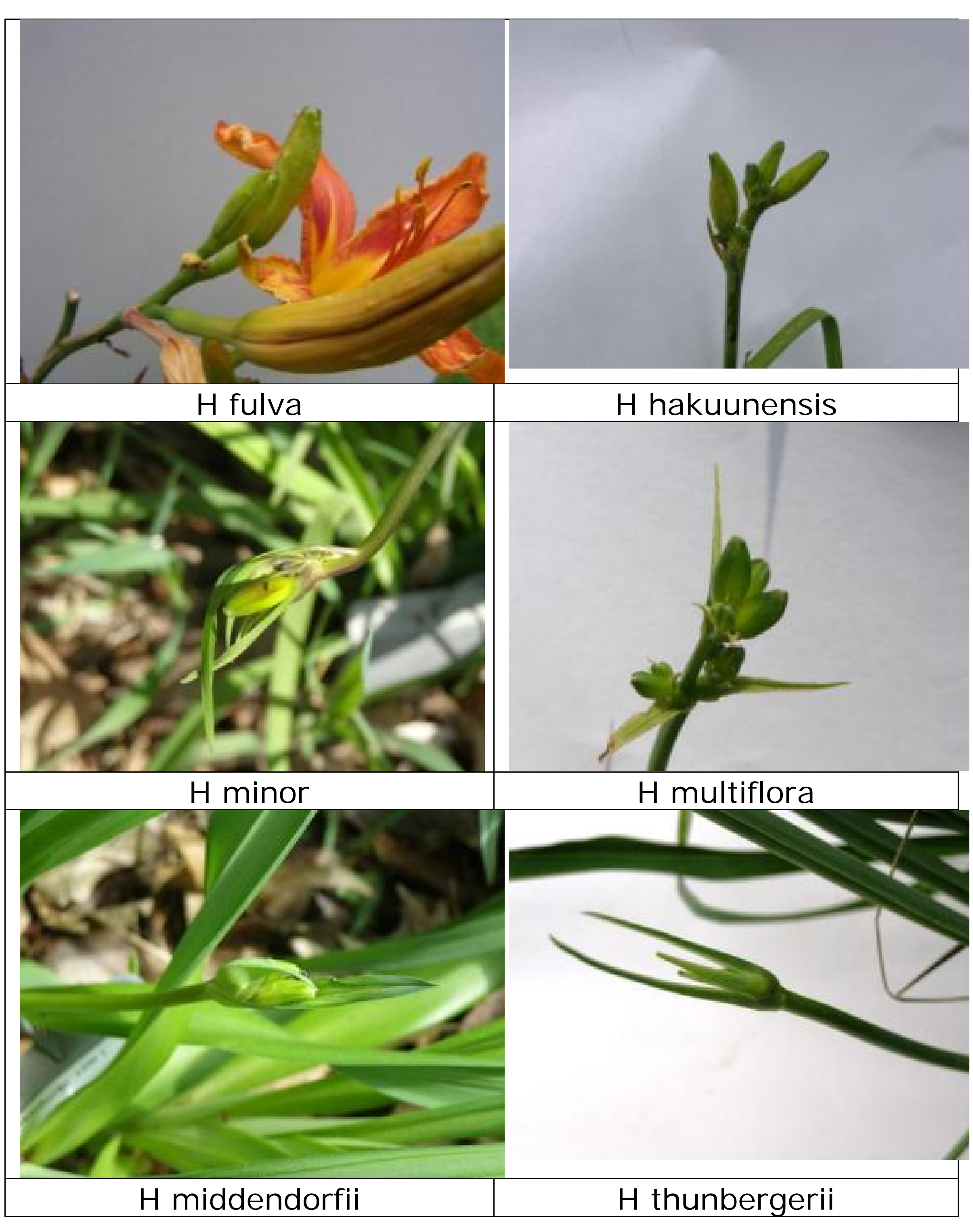


H fulva	H hakuunensis
H minor	H multiflora
H middendorfii	H thunbergerii

2.10.2 Buds

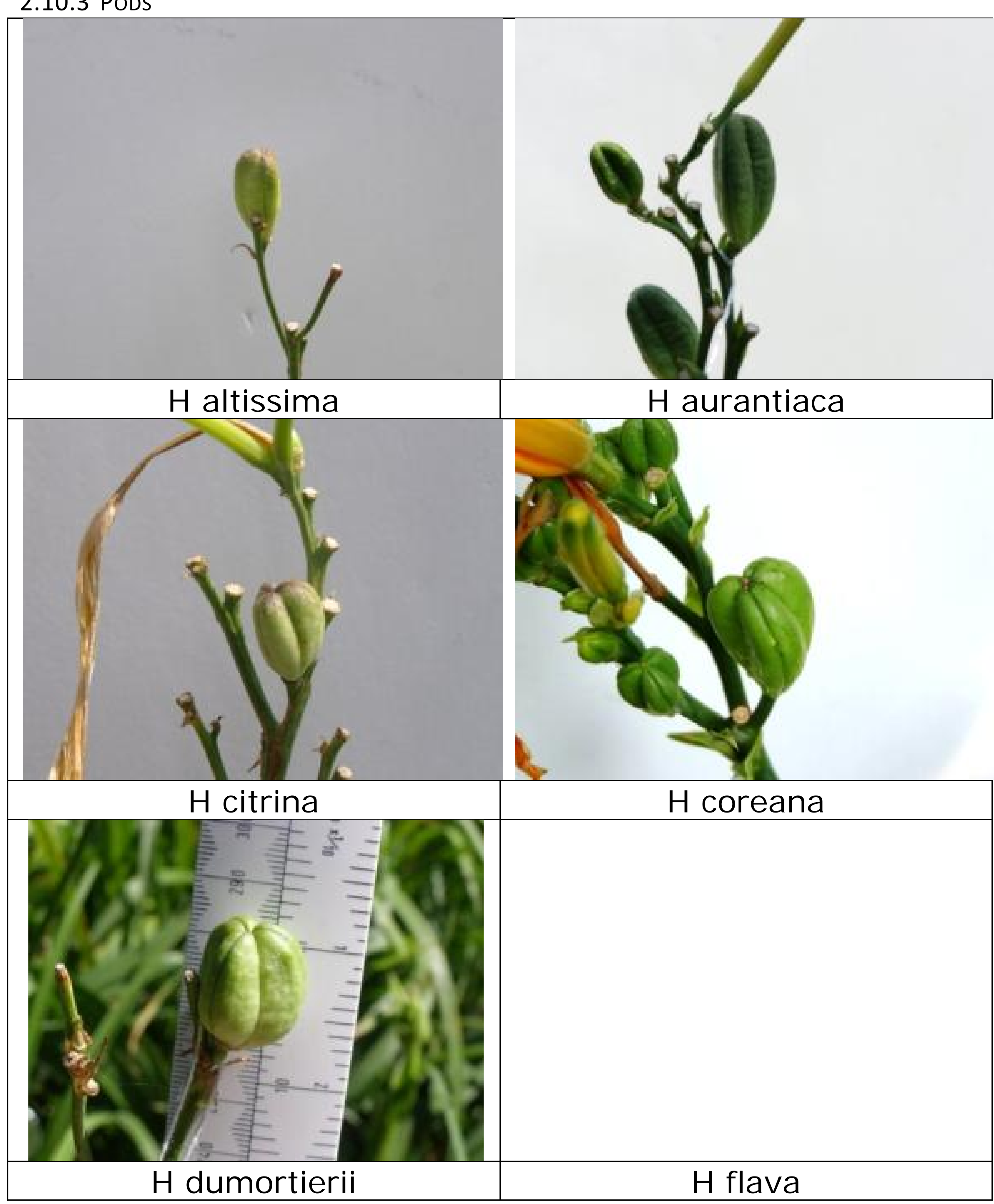


H altissima	H aurantiaca
H citrina	H coreana
H dumortierii	H flava

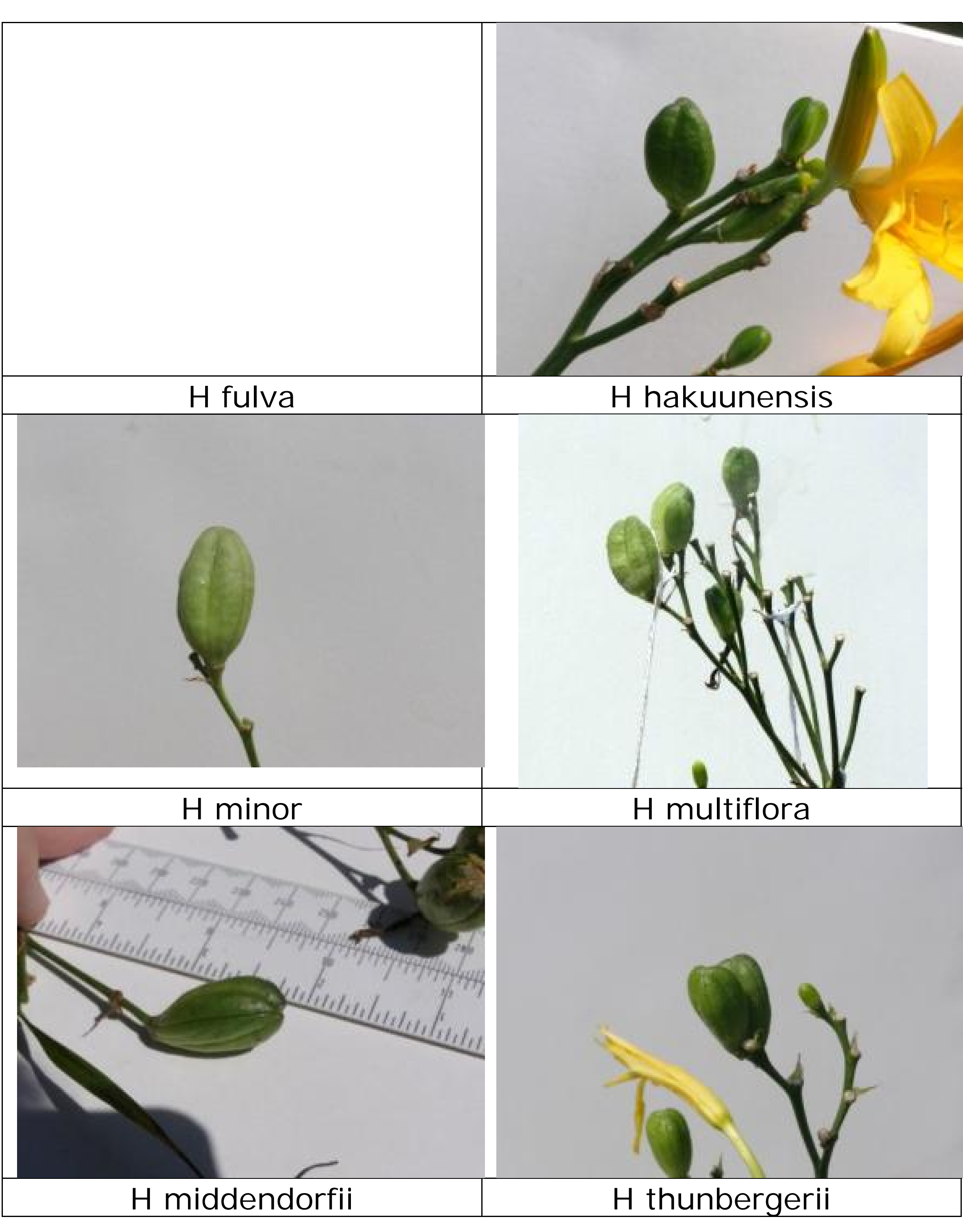


H fulva	H hakuunensis
H minor	H multiflora
H middendorfii	H thunbergerii

2.10.3 PODS

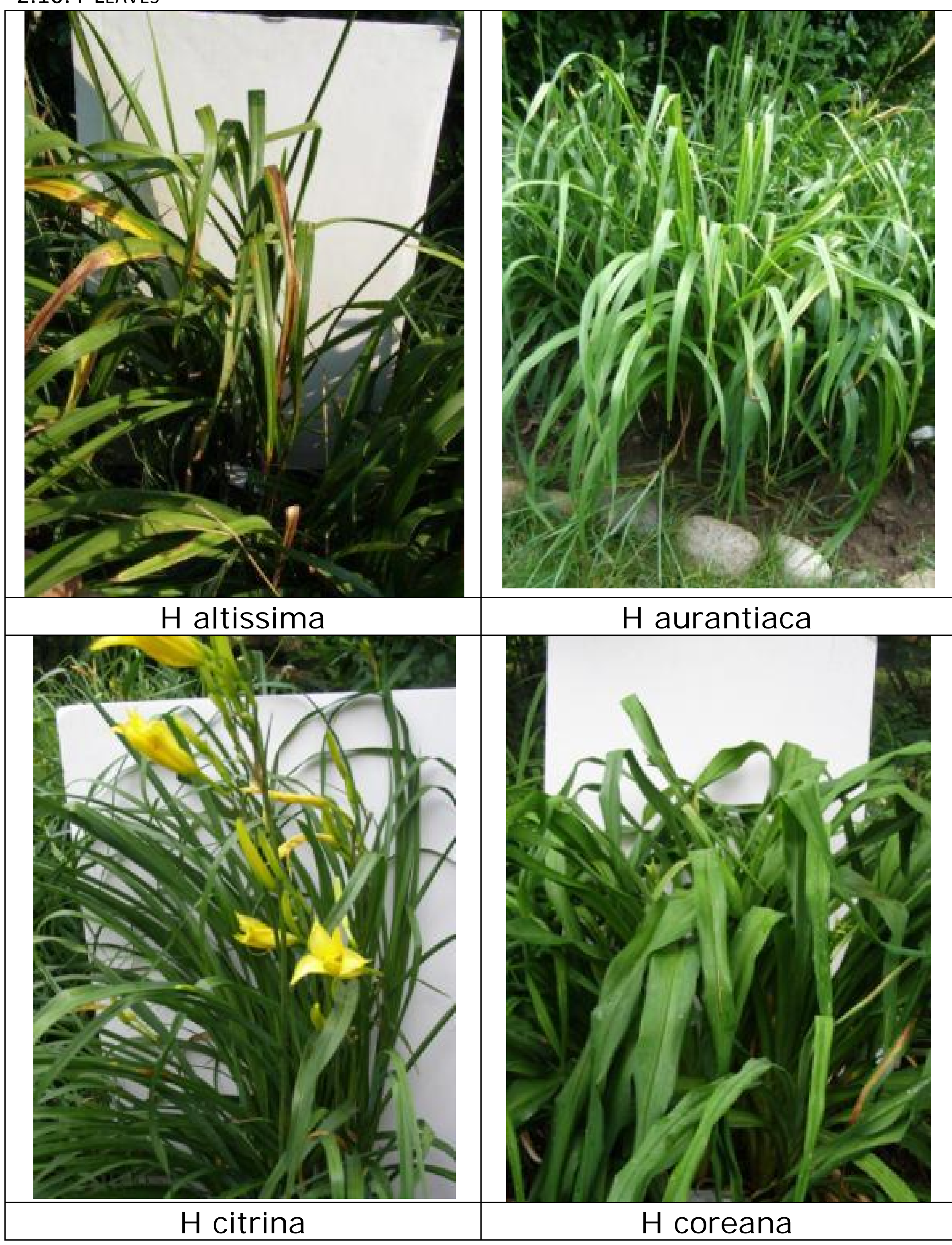


H altissima	H aurantiaca
H citrina	H coreana
H dumortierii	H flava



H fulva	H hakuunensis
H minor	H multiflora
H middendorfii	H thunbergerii

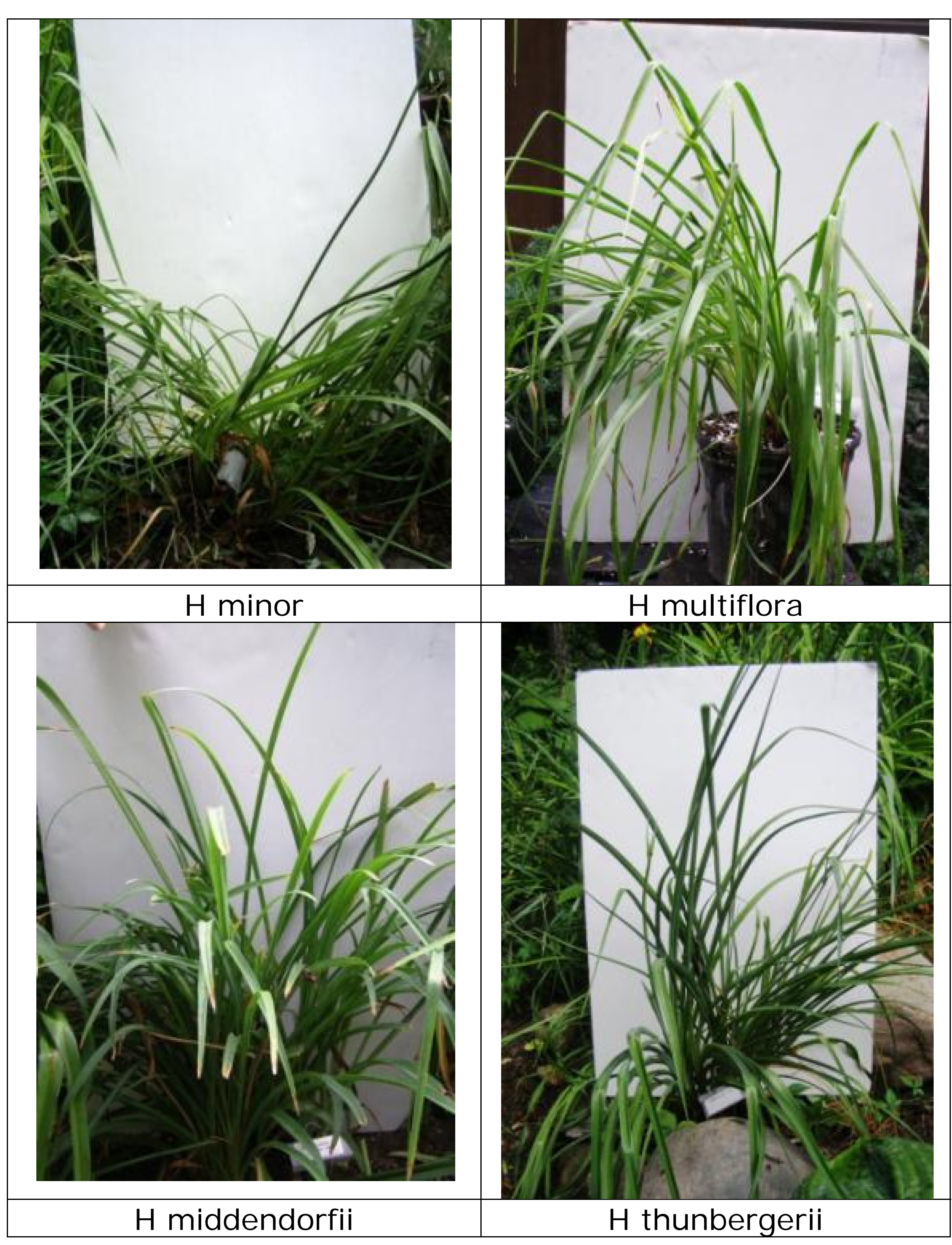
2.10.4 Leaves



H altissima	H aurantiaca
H citrina	H coreana

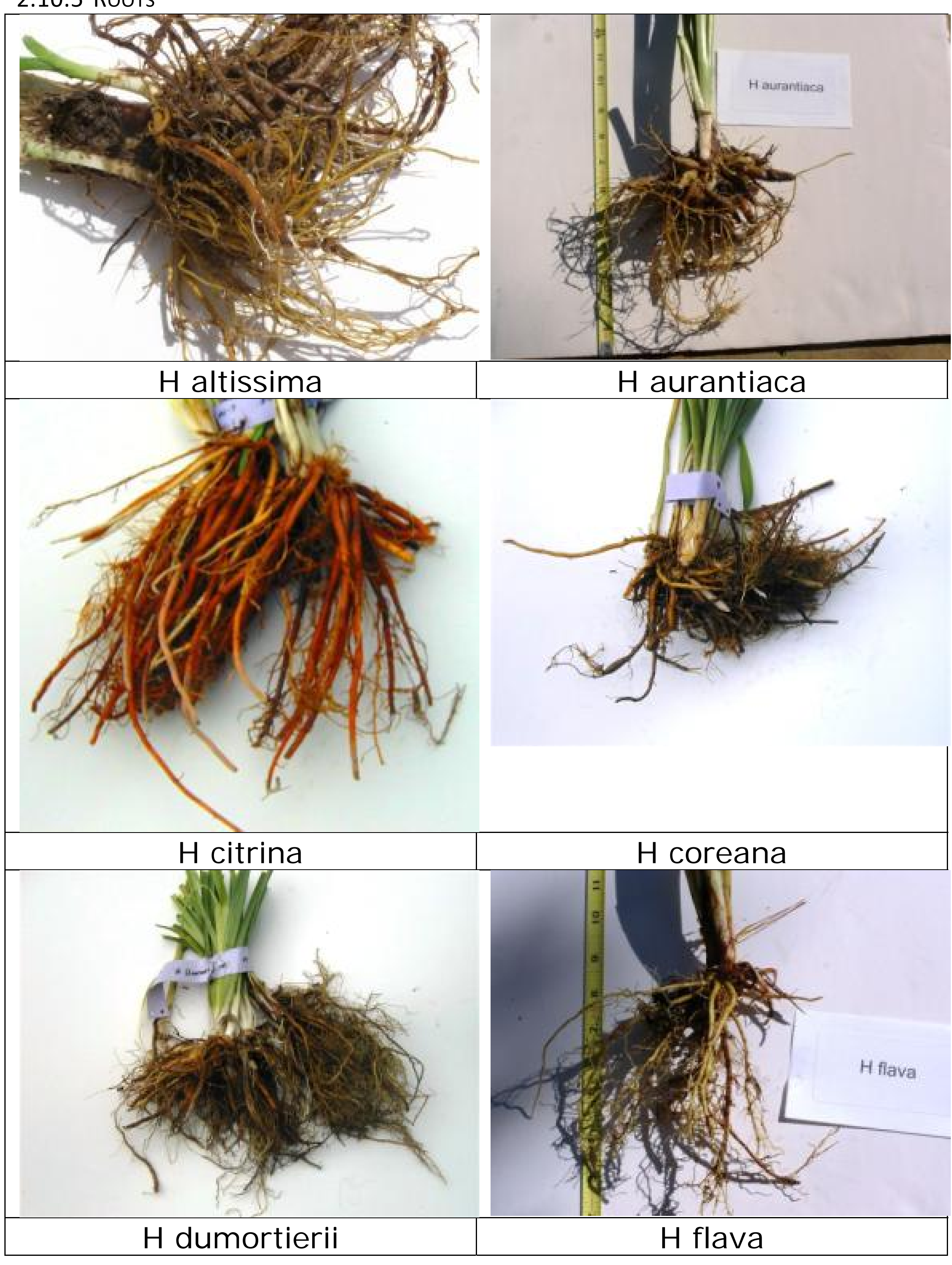
H dumortierii | H flava

H fulva | H hakuunensis

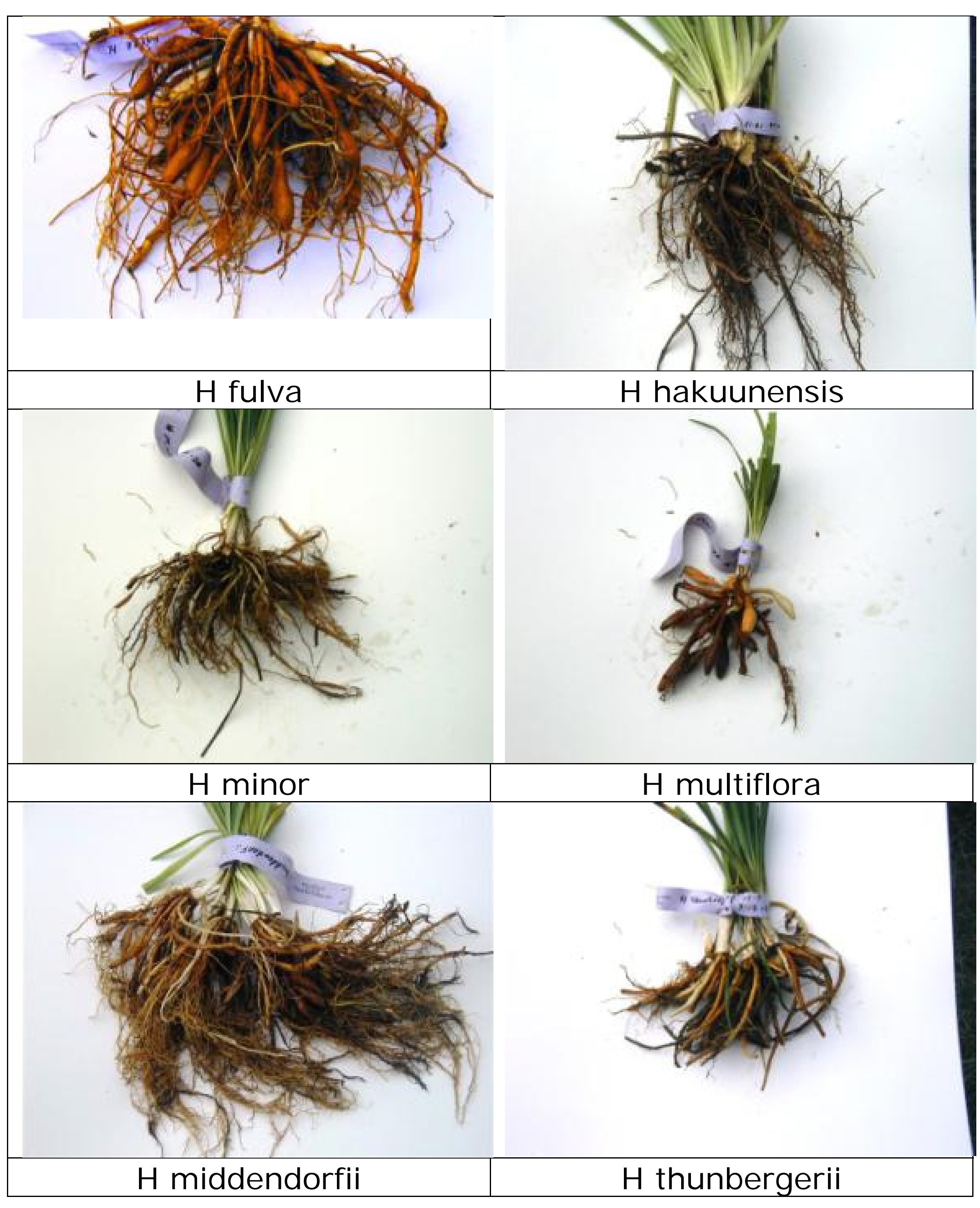


H minor	H multiflora
H middendorfii	H thunbergerii

2.10.5 Roots



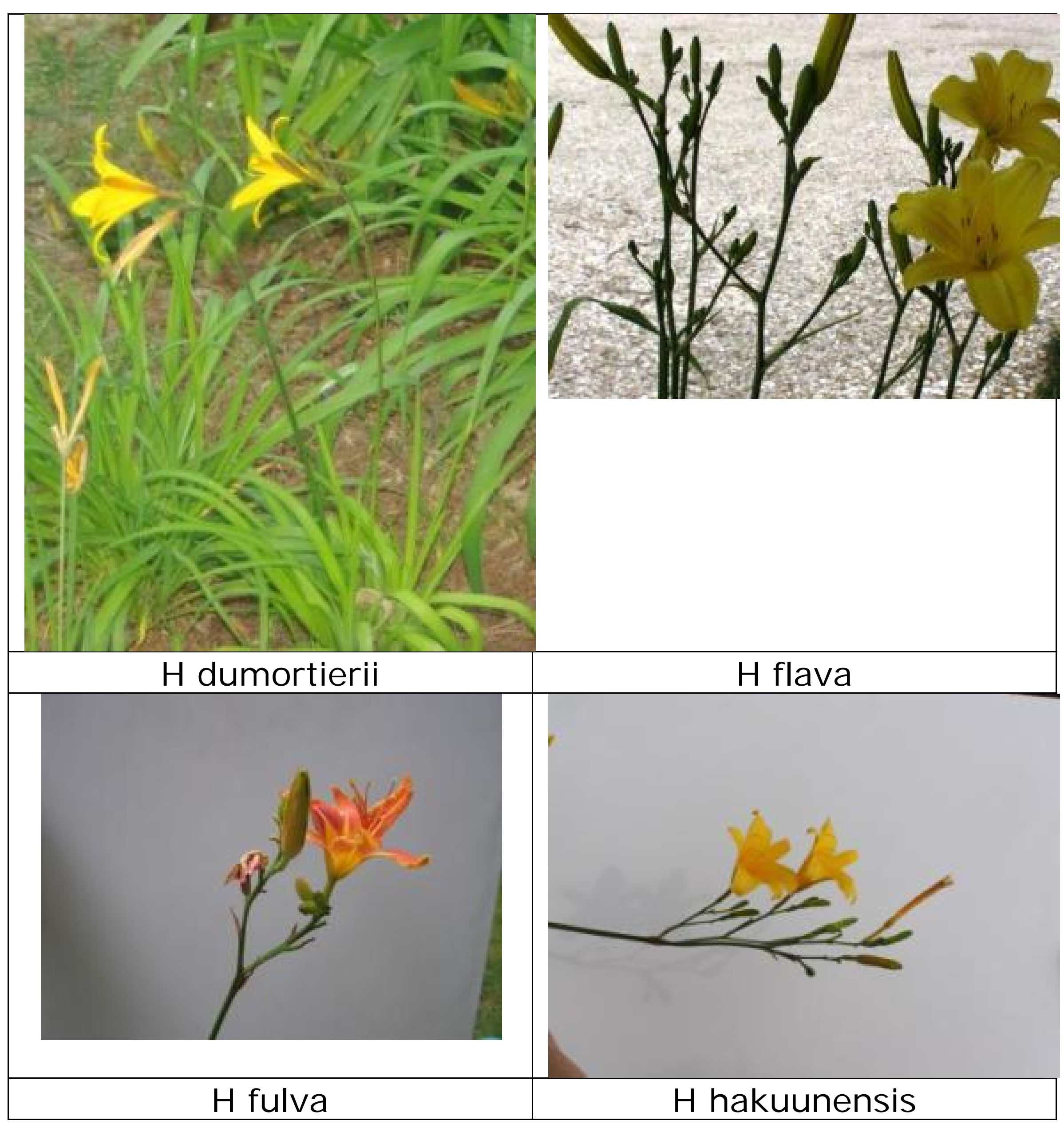
H altissima	H aurantiaca
H citrina	H coreana
H dumortierii	H flava



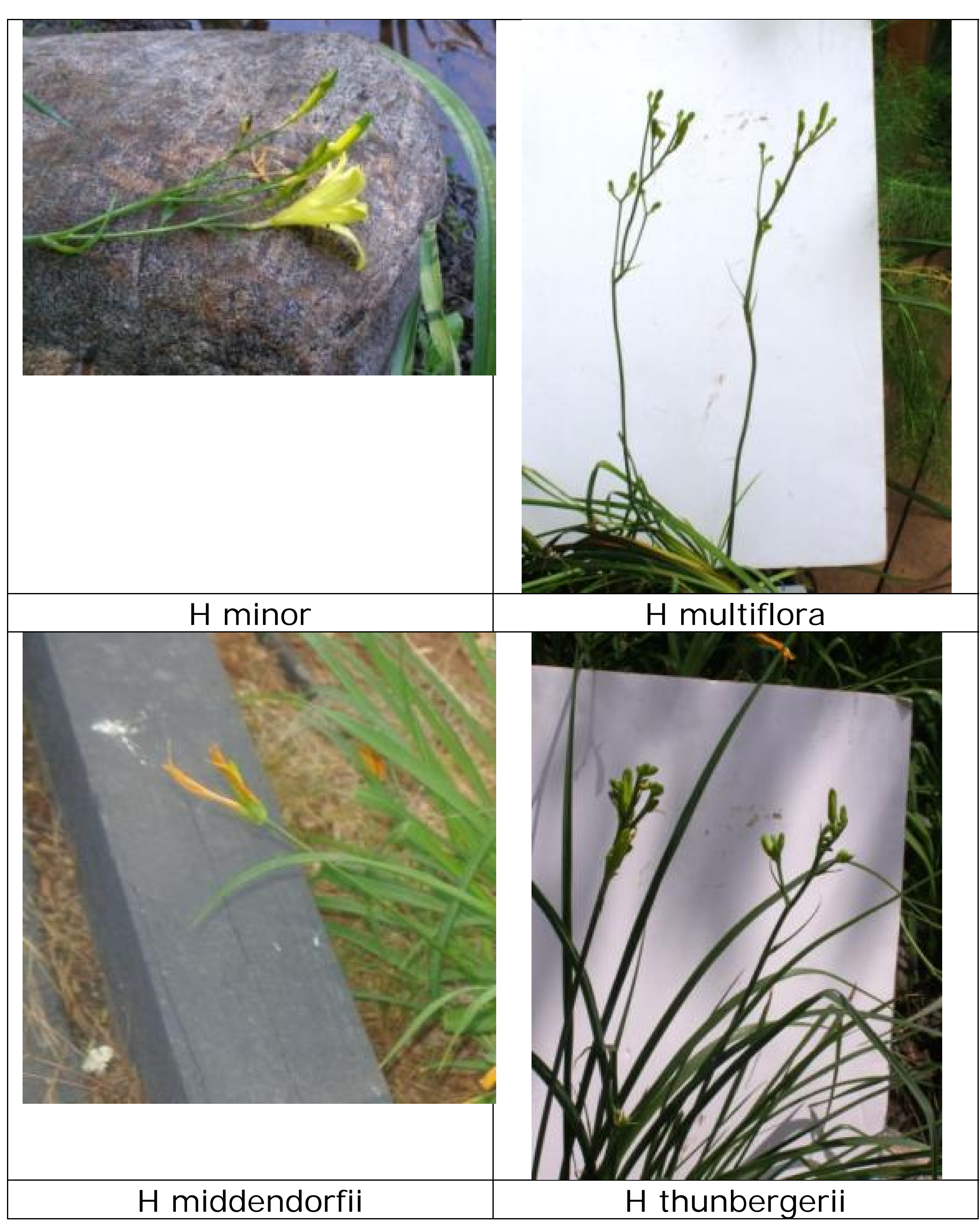
H fulva	H hakuunensis
H minor	H multiflora
H middendorfii	H thunbergerii

2.10.6 Branching

H altissima	H aurantiaca
H citrina	H coreana



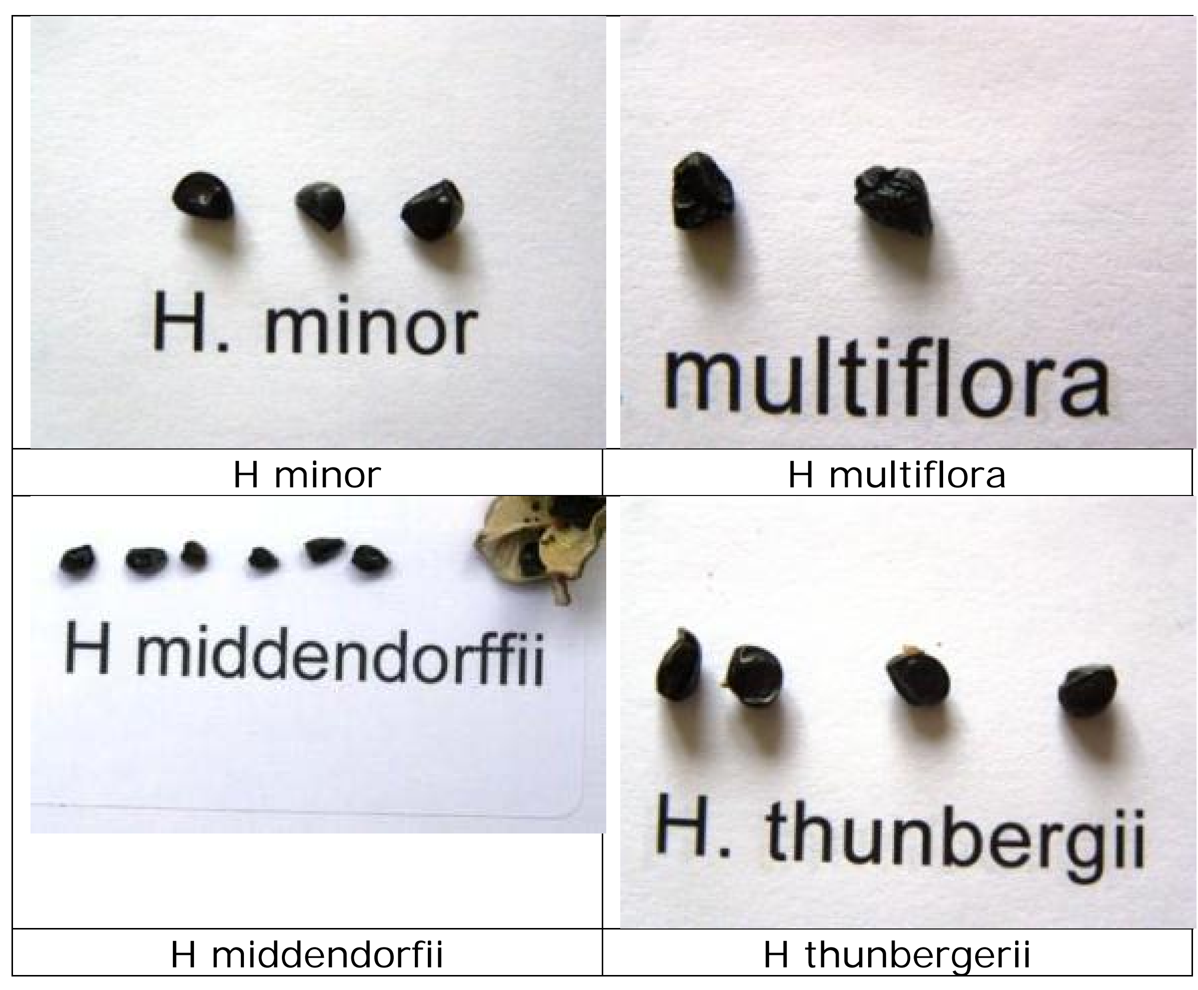
H dumortierii	H flava
H fulva	H hakuunensis



H minor	H multiflora
H middendorfii	H thunbergerii

2.10.7 SEEDS

H altissima	H aurantiaca
H citrina	H coreana
H dumortierii	H flava
H fulva	H hakuunensis



H minor	H multiflora
H middendorfii	H thunbergerii

Species	*Flower*	*Bud*	*Branching*	*Root*	*Pod*
H altissima					
H aurantiaca					
H citrina					
H coreana					

Species	*Flower*	*Bud*	*Branching*	*Root*	*Pod*
H dumortieri					
H flava					
H fulva					
H hakuensis					
H middendorfii					

Species	*Flower*	*Bud*	*Branching*	*Root*	*Pod*
H minor					
H multiflora					
H thurbergii					

3 GENETIC PRINCIPLES AND APPLICATIONS

This Chapter is a review Chapter for the subsequent document on the Genus Hemerocallis. This Chapter establishes the baseline facts as are currently experimentally know and which are at the heart of understanding the genetics of the Genus Hemerocallis. This Chapter develops models which will be used elsewhere in the analysis and synthesis of color and patterning of the various hybrids as well as establishing an understanding of the underlying sets of species and their resulting hybrids.

3.1 INTRODUCTION

The genetic structure of the genus Hemerocallis and its impact on the color and patterning requires an understanding of a few essential facts from the now well understood operations of the gene and the secondary pathways associated with them. This Chapter is a review of these principles. Specifically we review the following:

1. Gene structure and operation. This includes the basic Watson and Crick model as is currently understood. The development that we use is a functional model and note one that would be more familiar to the biologist. In all our analyses we will build models of functions and leave the basic principles and their modifications to the bench scientist.

2. Secondary Pathways are introduced and the related gene controls are presented. The secondary pathways which create the chemicals which in turn create colors are discussed in some detail.

This discussion should provide the basic principles to address the other issue we seek to develop.

3.2 PRELIMINARY CONCEPTS AND DEFINITIONS

We want first to develop some concepts and definitions. To fully understand the genus Hemerocallis and to be able to employ the techniques of breeding, one must have a common framework of concepts, the building blocks of the ideas we will develop and employ. This chapter begins that process.

There are several concepts we can begin to define. They are:

Chromosomes and Genes: The essence of understanding and growing new types of Hemerocallis is the understanding of the chromosome and gene. The Hemerocallis gene is somewhat simple and akin to that of a human. We humans have 22 chromosomes plus a sex chromosome. For a total of 23 chromosomes. The Hemerocallis has 11 chromosomes and no sex chromosome. Both generally have chromosomes in pairs, the human has 23 pairs of 46 chromosomes and the Hemerocallis has 22 pairs of 11 chromosomes[13].

Genotype and Phenotype: We all know what a specific plant "looks" like if we see it. We know its color, its size, its shape, and other characters or characteristics which we could then communicate in a somewhat unambiguous fashion to others so that they could in turn say whether they have found the same "type" of plant. In contrast, we can now ascertain the genotype of a plant, at least on the small. We can look at certain gene loci and from them determine what the plant is. In today's world we can use this genetic information perform various analyses which in turn will allow us to "characterize" a specific plant. But we know that no two plants have exactly all the same genes, some genes may not be expressed, so they may "look" alike in all aspects, but hidden in sections of their DNA are segments which do not speak but are different.

Species: Just what is a species and what does it mean for us as we proceed through this study. This is the most critical question that we shall pose and we shall spend considerable time discussing its meaning.

Breeding versus Hybridizing: Daylily people consider themselves hybridizers. Agricultural botanists look at breeding, as do say people who raise dogs, horses and the like. What is the difference between breeding and hybridizing and which applies or should apply in this area.

Pollination, Self and Cross: Obtaining variability in a plant means we must work with what is in nature or has already been developed by others. The plant in the wild will pollinate itself or will cross pollinate with others. What do we really mean by these terms and how does that influence the concepts we are trying to develop.

3.2.1 Chromosomes and Genes

Let is start with the chromosome. We will return in some detail to this latter but at this point we want to establish a few basic definitions. The plant has 11 pairs of chromosomes, for a total of 22 chromosomes[14].

[13] See Kang and Chung (1997) p. 210, Journal of Plant Research, Japan. The authors state that they have independently verified this number.
[14] See Kang and Chung, 1997, Journal of Plant Research.

The Figure below is a graphic of a typical plant cell showing the nucleus and one of the chromosome pairs. This graphic is not at all what one would see in reality but it is typical of the generic elements.

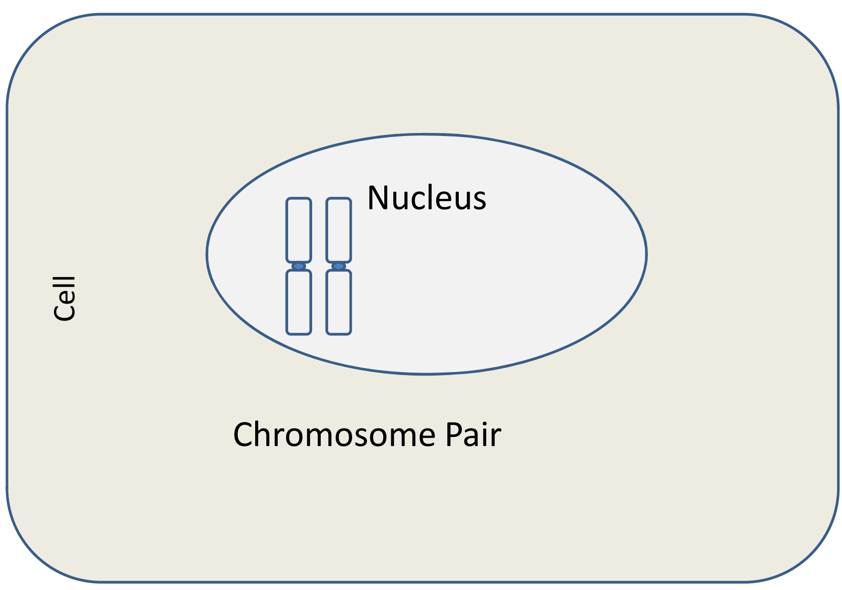


FIGURE 60 BASIC CONSTRUCT OF A PLANT CELL.

3.2.2 CHROMOSOME

The chromosomes are the collection of DNA which agglomerates together into separate units. They bind together as pairs and it is these pairs which make up the chromosomes we see in the nucleus of a mature cell.

The Figure below depicts the types of possible chromosome combinations we would see in a typical Hemerocallis. This is called ploidy, haploid being one chromosome and diploids being pairs of chromosomes.

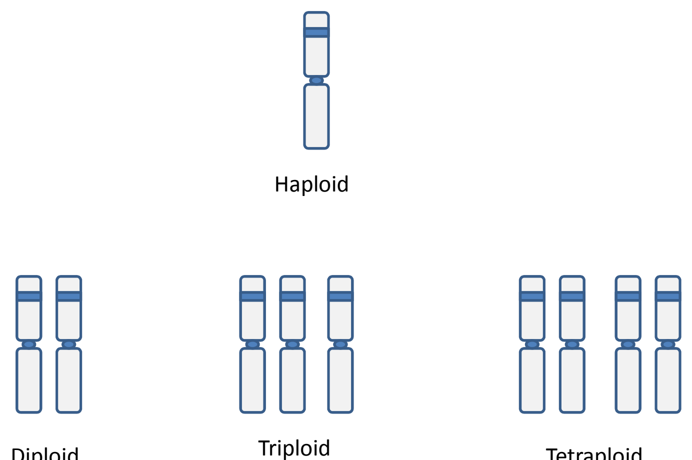


FIGURE 61 PLANT PLOIDY

The types of ploidy are:

Haploid: The haploid is the single chromosome strand that one may be able to see in the sex cells of a plant. Namely in the pollen or in the ovary cells. The haploid is a single stranded non-binded collection of DNA.

Diploid: The diploid is the prototypical collection of DNA in the mature Hemerocallis as is normally found in species and in many hybrids. The diploid is merely two, one from the male and one from the female.

Triploid: This type of three way bonding is found in many Hemerocallis which do not produce sexually such as the H. fulva Europa, the common garden variety orange daylily and the doubles we see frequently the H fulva Kwanso and H fulva Flore Pleno. These triploids are not at all readily used in crossing but it has been recorded that from time to time they do manage a cross. The details of the crossing mechanism are not fully understood.

Tetraploid: Since the mid 20th century, with the use of colchicine an alkaloid from the Colchinum genus, also used for gout, the creation of tetraploids was possible. Tetraploids have four chromosomes per grouping and thus the nucleus has a total of 44 chromosomes. This is twice the DNA of the normal diploid and this doubling introduces many additional variations which we shall show later.

3.2.3 DNA

DNA, deoxyribonucleic acid is the heart of the gene. It is the basis of the code we can understand to determine the relationship between genes and their phenotypic responses.

We briefly layout the ideas concerning DNA in this section. DNA is constructed in the following manner. There are four base elements; Adenine (A), Guanine (G), Thymine (T) and Cytosine (C). They are shown below.

NH_2 N N N N H	Adenine
O N NH N H N NH_2	Guanine
NH_2 N N H O	Cytosine
O NH N H O	Thymine

FIGURE 62 BASES OF DNA ELEMENTS

These Base elements can combine in only a specific manner, namely A with T and G with C. These bonds are shown below. This was one of the seminal observations which drove Watson and Crick towards their great discovery. The bonding also is the basis for how these Bases combine in pairs, the Base Pairs, and then how these Base Pairs link up to form the now famous DNA chain.

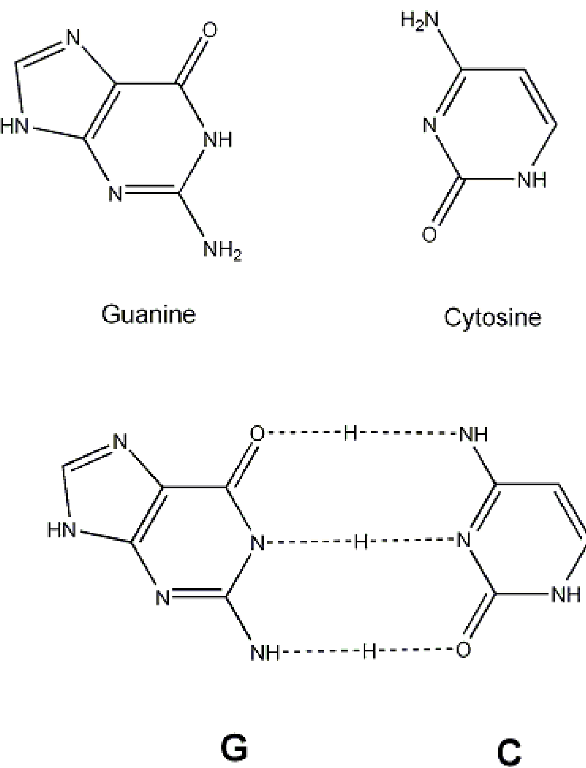


FIGURE 63 CT BASE ELEMENTS AND THEIR BONDING.

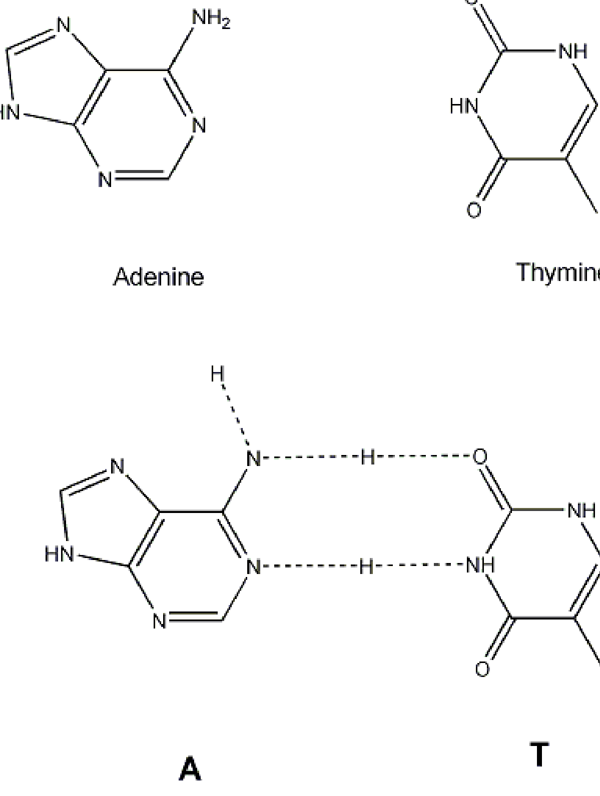


FIGURE 64 A-T BASE PAIRING

Now these Base Pairs are connected to sugar molecules, a cyclic ribose, to create a Nucleoside, such as deoxyadenosine. Then the nucleosides are enhanced with a phosphate constellation, a phosphorous molecules surrounded by oxygen and hydrogen. This combination of the nucleoside and the phosphate is called a Nucleotide. It is these nucleotides which connect on a backbone on the outside and in another backbone on the inside to form the DNA molecule. The following Figure shows a Nucleotide connection, we do not show the base pair connections. The Nucleotide has two defined ends; a 3' end which of the OH molecule and the 5" end which is the phosphate. We show these in the following Figure. These ends will play an important part in the generation of the products of DNA.

5' End

3' End

3" End

5' End

FIGURE 65 DNA NUCLEOTIDES AND 3' AND 5' ENDS.

The nucleotides are then connected into the long DNA wrapped double helix which is generally well known. This is shown below. Our interest will be in the genes themselves and we will look at them in some detail. One of the key questions will be just what is a gene? That will be a challenging question. It will go to the heart of hybridizing. It can be answered in many ways but clearly the simple ideas of Mendel must be revisited.

In the Figure below we set forth a paradigm of the opposite bases and they are lined up in a stretched out set of nucleotides where we are looking solely at the base elements, the A, T, G and C.

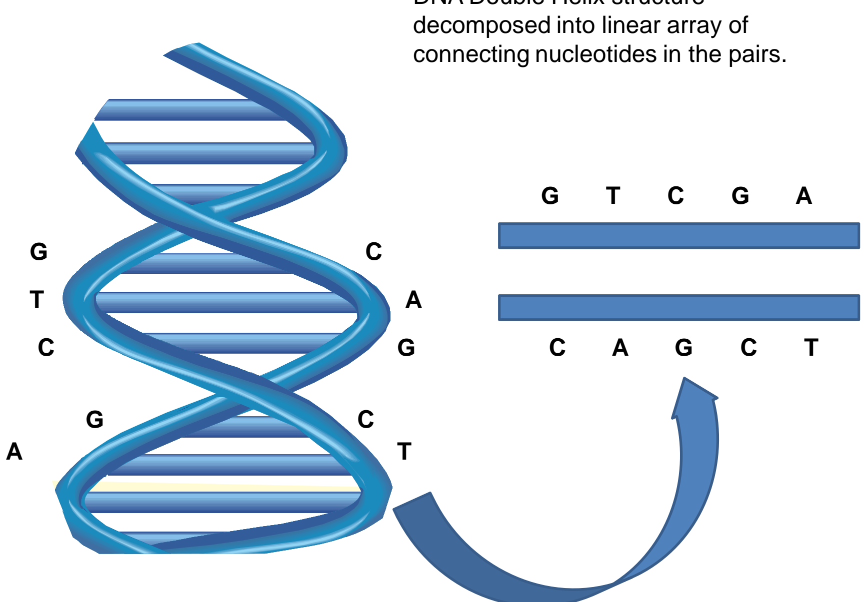
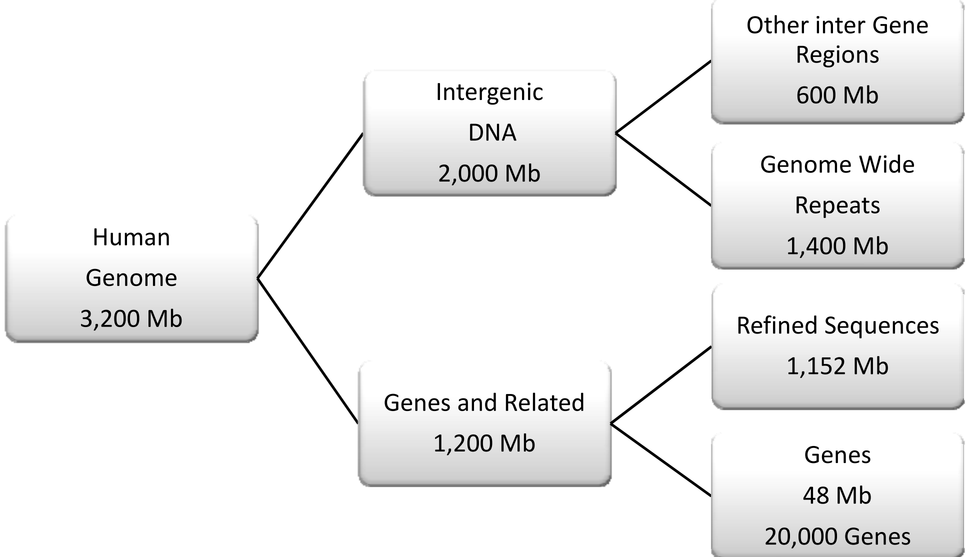


FIGURE 66 DNA DOUBLE HELIX AND BASE PAIR PARADIGM.

In the human the DNA is of moderate size, about 3,200 Mb, that is 3.3 billion G, T, A, or C. However as shown below the DNA is broken down into many small elements. The actual operating genes constitute a mere 48 million bases and this constitute about 20,000 genes. That is an average of 2,400 bases per gene. As we shall see it takes three bases to create one chemical compound on a protein, this there are a total of 800 per protein on average.

The main conclusion is that there is a great deal of what has been called junk DNA. That DNA is useful for identifying people, namely that is used in DNA identification, and it may or may not play great roles in protein generation and gene modulation.



Ref: Watson et al, Molec. Bio Gene 5th Ed, p. 137

3.2.4 Gene

The gene is the fundamental building block of any living creature. It is not the single expressive element to control a phenotype, it may contribute to that control but it is not the one to one element in the process. Thus a red flower may be controlled by several genes and in addition those genes may be affected by several epi-genetic factors ranging from the environment to other genes.

The human is now thought to have about 20,488 genes[15]. Not a large number and greatly lower than what literally all the experts thought before the human gene was fully analyzed. Many experts had guessed that there were well above 300,000 genes in the human. The Human genome is composed of slightly more than 3 Billion base pairs, combinations of G, T, C or A. The Hemerocallis genome is approximately 4 Billion base pairs. The number of active genes in Hemerocallis is at this time unknown. But it is close in size to the human genome.

The simple construct of a gene is shown below. It is a collection of DNA bases which combine together in terms of the effect. We show in the Figure the Introns, namely the unused DNA bases, and the exons, the used DNA bases. The exons are "combined" to effect what a gene does.

---

[15] See Pennisi, Working the (Gene Count) Numbers: Finally, a Firm Answer? SCIENCE Vol 316 25 May 2007 http://www.sciencemag.org/cgi/content/full/316/5828/1113a?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&fulltext=gene+count&searchid=1&FIRSTINDEX=0&resourcetype=HWCIT

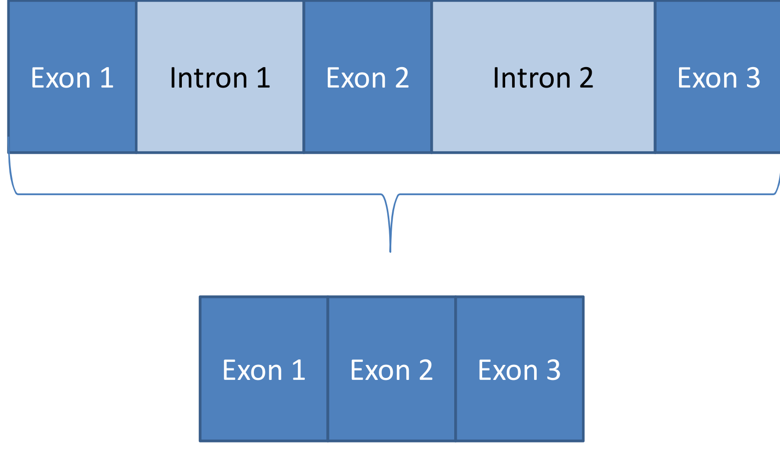


FIGURE 67 GENES AND INTRONS AND EXONS

What then is a gene? For our purposes and to be consistent with contemporary understanding we define a gene as:

"A gene is a collection of DNA bases which when combined in a determinable manner can express the combination of bases via the production of some effect upon the cell and potentially its surrounding environment. A gene is an expressible collection of base pairs, when acting in concert, in the internal environment of a cell."

Thus we understand a gene by its effects, not just by its structure. It effects may be complex. It may produce some RNA, and in turn a protein, it may activate or suppress another gene, or it may be the basis for creating a new gene in this construct. Based upon what we know and understand today, a gene is not some well defined coherent set of contiguous DNA. Genes can even be created on the fly within the cell based upon the environment that is if we define a gene by what it creates and affects.

The classic paradigm for DNA influence is shown below. Namely that DNA generates RNA via transcription and RNA generates proteins via translation. We will not get into further details other than saying that this process has many sub elements which will be regarded in further detail latter.

DNA

DAN are Exons combined and controlled by RNA

RNA

DNA Transcribes RNA

Protein

RNA Translates into Proteins

FIGURE 68 CLASSIC DNA PARADIGM

The above understanding of the gene and its relationship to its environment states that there exists a gene, a construct, which uniquely generate an RNA strand, which in turn uniquely generates a protein. We now know that these are all subject to further analysis. For example, the gene is not just a connected set of DNA bases, it is a set of exons, which may be combined in a sequence, or may even be broken or reassembles. Thus the gene is determined by what it does, not by any unique set of base pairs.

The protein that results from the above model is then related to some phenotypic response.

3.3 Genotype and Phenotype

Phenotypes are what we see, smell, hear, touch, taste; they are the interactions between some creatures, in our case a plant, which we may use to identify the plant. In the genus Hemerocallis, the phenotype may be the color, color patterns, size, time of bloom, odor, texture of the flower, and other definable characteristics that we see when we observe the plant.

Genotype is what the gene has as specific content, its specific DNA. The production of a phenotype is frequently driven by the expression of a gene. The gene "expresses" itself in a very special manner. The DNA is wrapped in tight coils.

The model we will build upon appears as in the Figure below. This is the canonical model for gene expression. We assume that there is some collection of secondary pathways, and that these pathways result in chemical products that are directly related to a phenotype; a darker red flower, a longer leaf, a taller scape. That these pathways are

modulated in some manner by proteins generated from within a cell. That the proteins are the result of some entity called a gene. That the gene can be an assembly of bases and the gene may itself be modulated up or down by activator or repressor proteins respectively generated by other genes or even the same gene. Thus we model the cell as a dynamic system and further we argue that this system has certain random elements which we shall include latter.

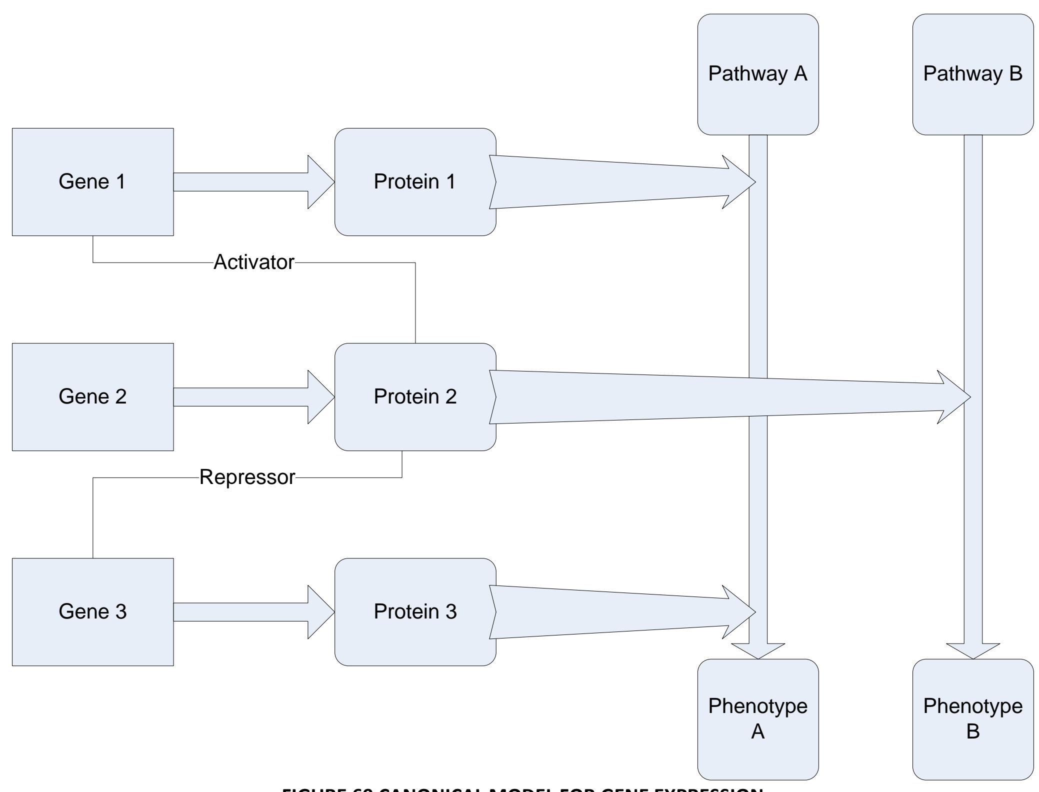


FIGURE 69 CANONICAL MODEL FOR GENE EXPRESSION

It is the output of this genetic process that we get the plant in its full temporal and spatial existence.

The above model of the gene is one in which we see the beginnings of some form of feedback. We see the activator and repressor genes as the basis for this element. However this may be expanded even further, We show this below. Note we show that the Gene K can be influenced by other Genes, as well as the products of the pathways as well as by the environment. The Environment can modulate the pathway which by being fed back to another controlling gene can then modulate the activating gene. This process is a complex process and exceeds what we would have imagined from the simple Mendellian gene theory.

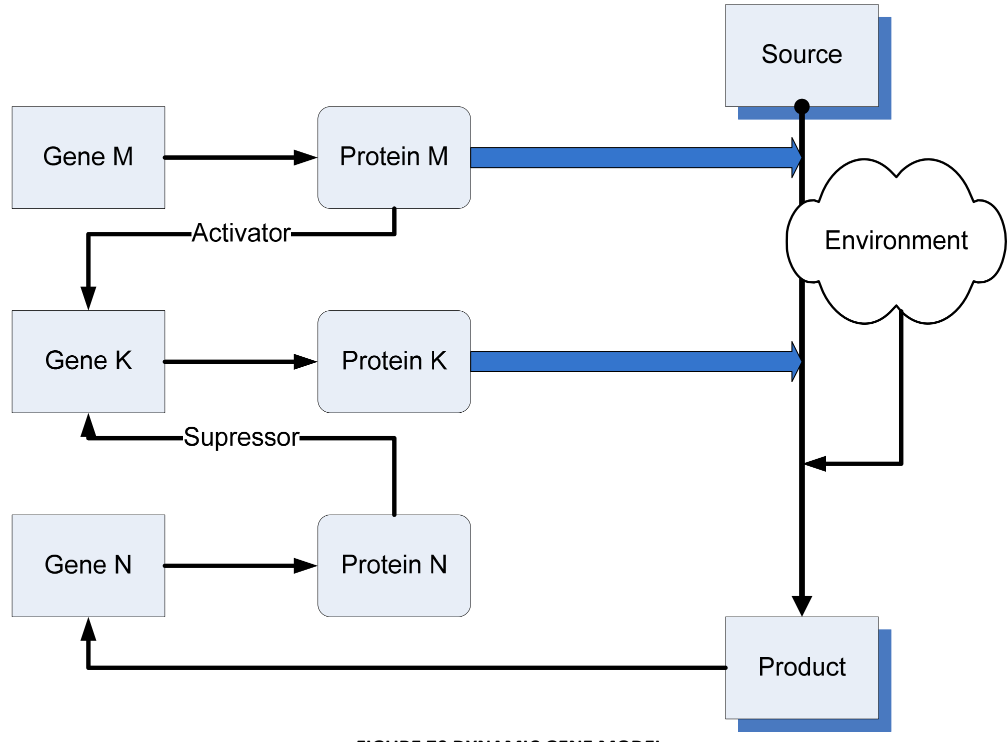
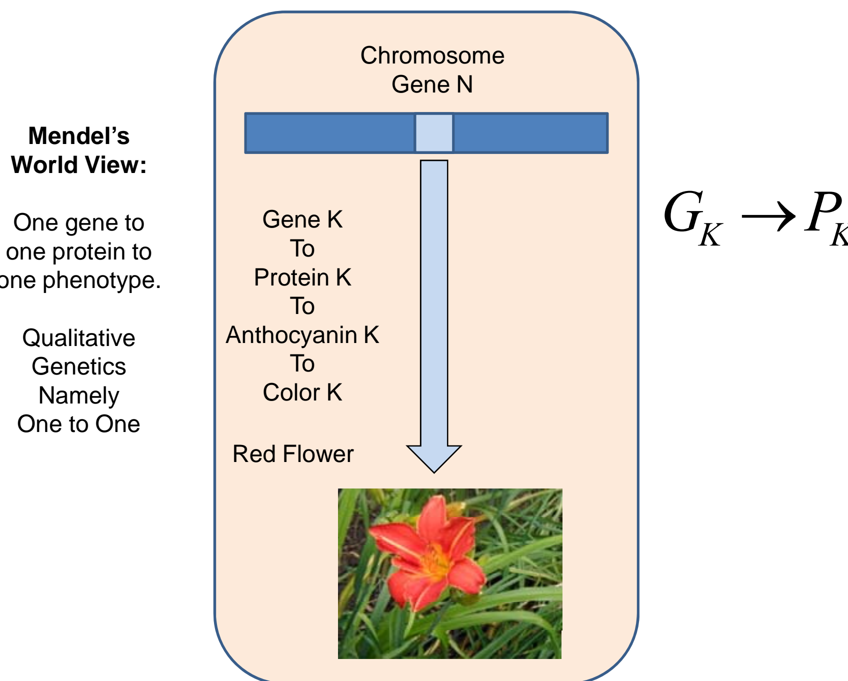


FIGURE 70 DYNAMIC GENE MODEL

Now back one again to the Mendellian Gene model. Although Mendel and his model was not so rigidly simple, for he did admit some other influences as well as variation, we will call the simple Gene and Phenotype combination the Mendel Model. Namely in this model we assume the existence of a Gene and then we further assume that there is some phenotypic characteristic such as flower color which maps one to one onto this gene. One gene and one phenotypic character. The phenotypic characters further have countable and discrete values. The flower is red, yellow, and green. There are no blends and there are a limited numbers. Then there is a gene for red, a gene for yellow and a gene for green. The gene is at the same place on the chromosome and the gene just somehow changes to produce a different color. In addition the genes are dominant in some order. That is if there is one red gene, of the two on the chromosome, then we get red, if not a red but a yellow we get yellow, and we get green if and only if there are all green genes, namely two.



Now there is a second model, based upon our understanding of DNA and the Watson Crick world. However this model goes well beyond the simple Watson Crick model. Here we assume we have long segments of DNA with many exons and many more introns. The gene as we know it is the result of the cellular processes which assemble the exons into a block of DNA which RNA will use to in turn generate a protein. In reality what happens is that the exons may be recombined to generate RNA in a variety of fashions. The result of that process, as well as the dynamic model we depicted above is that the phenotypic characteristic, say leaf length or width, or date of first bloom, takes on the character of a random variable. It has a set of values whose probability distribution may be of some form. We use as an example a standard Gaussian curve. This is shown below.

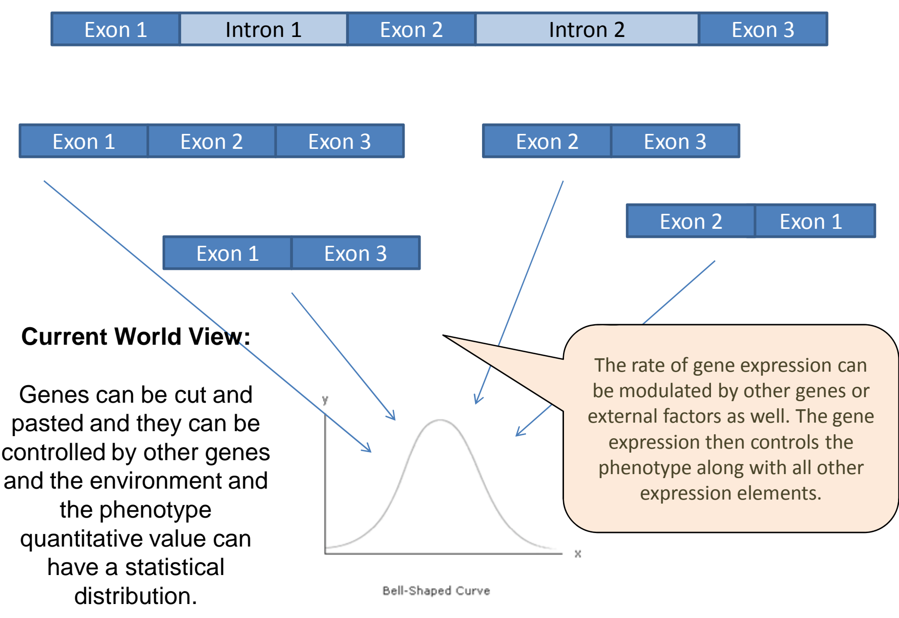


FIGURE 71 CURRENT VIEW OF GENETIC CONTROL

For example we show in the following Table results from Hasegawa et al as modified:

Table 1. Morphological traits (mean and standard deviation) of *Hemerocallis citrina* and *Hemerocallis fulva*, their F1 hybrids, and individuals in the hybrid population The standard deviation is given in parentheses

	H fulva	H citrina	F1
No. of scapes	72	74	55
Flower tube length (mm)	32.40	45.28	32.27
Petal length (mm)	92.51	81.13	78.60
Petal width (mm)	15.42	14.50	16.01
Stamen length (mm)	77.02	68.04	64.39
Pistil length (mm)	97.34	76.19	77.73

We now use the data from Hasegawa and present their curves shown the statistical distribution of scape and dimension.

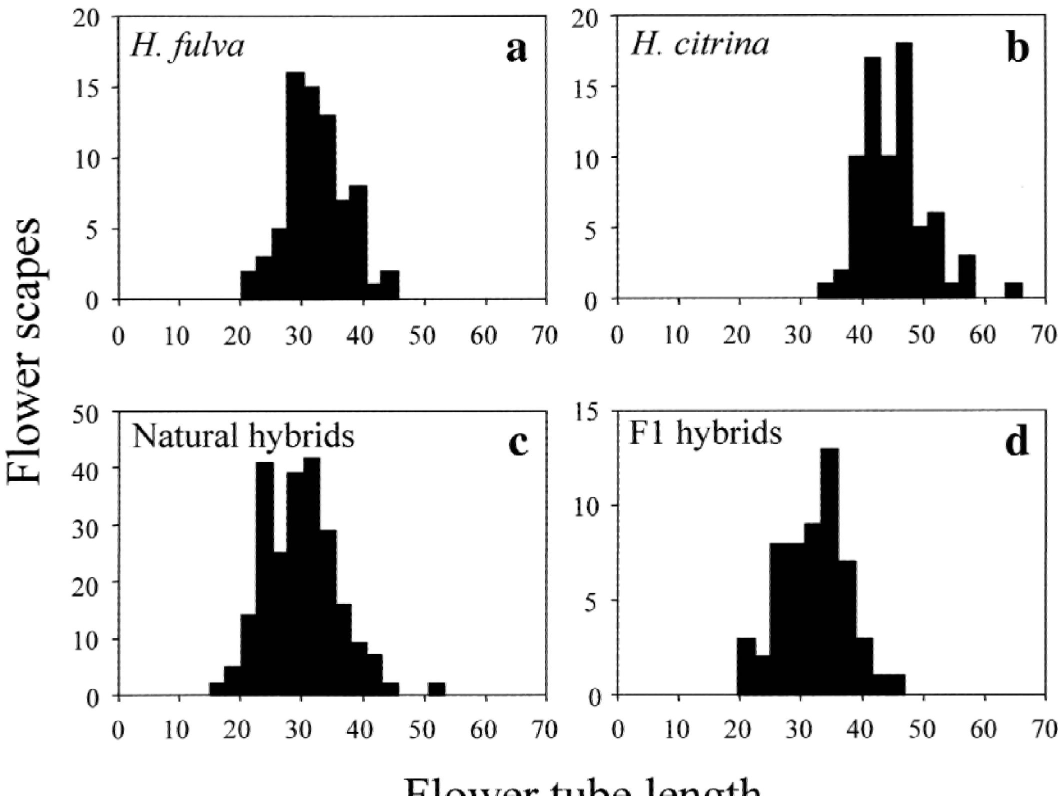


FIGURE 72 FROM HASEGAWA ET AL, DISTRIBUTION OF H FULVA AND H CITRINA

In a further step using Hasegawa and their study of H fulva and H citrina we can see the distribution in flowering time of the two species as recorded by the authors. Clearly several things can be observed. First, the time of bloom is clear bimodal and this form of separation will enhance the separateness of the species. The pollinators further reinforce the separation. Second, and in line with our above discussion, the blooming is spread out over a large period of time. This gives a probability distribution, albeit not a Gaussian one,

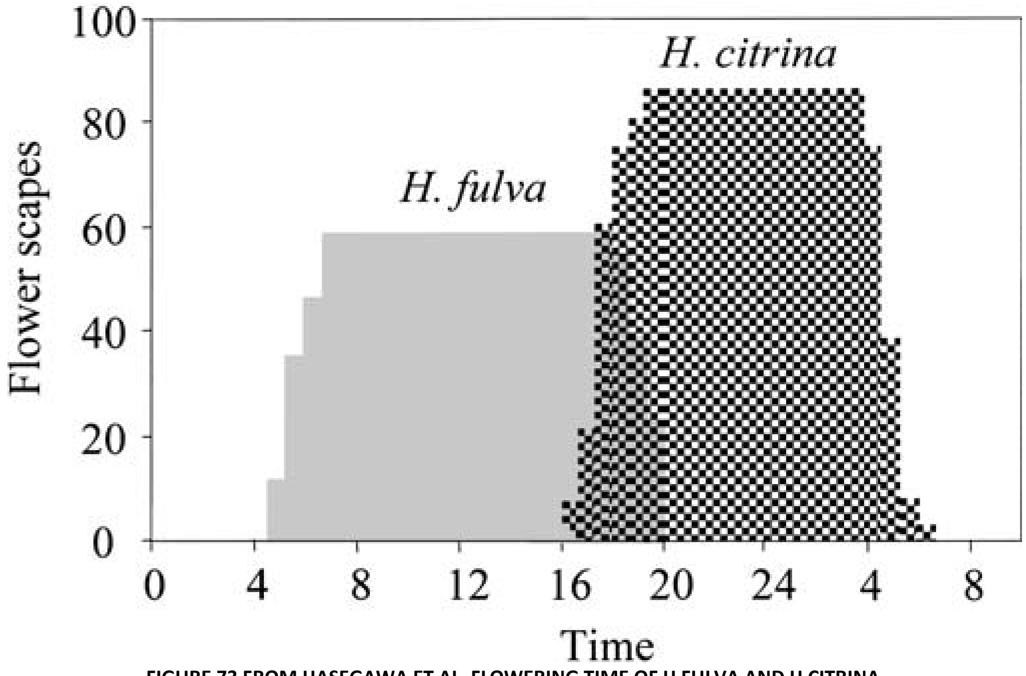
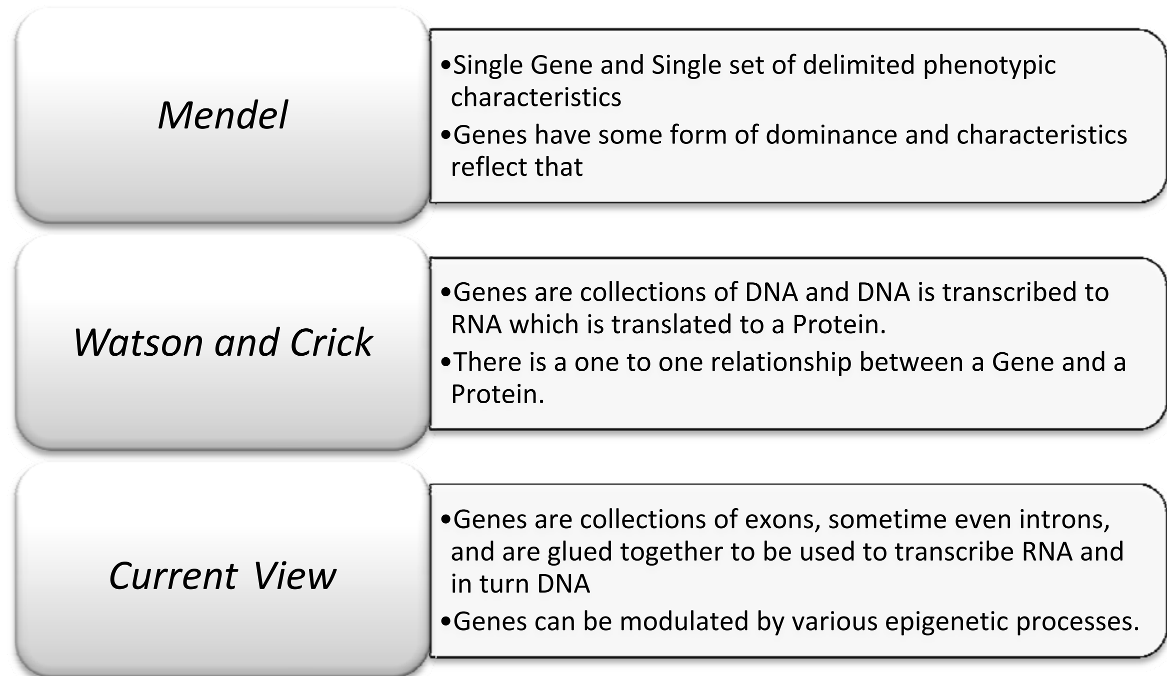


FIGURE 73 FROM HASEGAWA ET AL, FLOWERING TIME OF H FULVA AND H CITRINA

The views of gene impact are summarized in the following Figure. We shall use this model as a way to better understand how one can better seek hybridizing opportunities.



3.4 Genetics

In this section we present an overview of the classic Mendellian analysis.[16] The Mendellian analysis makes classic assumptions which prevailed until the advent of the Watson and Crick model, and even slightly beyond. In fact many breeding programs build upon a Mendellian approach. We argue that such an approach is partially correct but lacks most of the key elements which must be considered.

In this section we briefly review the molecular genetics of a plant cell. We do not get into any significant details but merely review the elements which we can use letter in developing the mathematical models for plant regulation. As we have shown in the previous section, plant colors are the result of the expression of three types of secondary plant cell products; anthocyanins, flavones and carotenoids. We have focused mainly on the anthocyanins but have shown the details on all three. What we focused on is that the production of any one of these is a result of a specific pathway and that the production in that pathway is controlled by a set of enzymes. The enzymes are proteins produced within the cell. The proteins are the result of the expression of a set of genes.

In this section we now by reviewing the current understanding of plant cell micro genetics show that the proteins are expressed by the normal process understood since Watson and Crick's seminal work and that there are factors which and activate their production, indeed enhance their production, or repress their production. These are the activators or repressor proteins. The activator and repressor proteins are in effect other genes expressing themselves. We will combine the last section with the results in this section to affect a dynamic system model for plant color generation in the next section.

What will be critical to understand here is that we just want to place the process of activators and repressors in context. We discuss in the next section what our overall design approach will be; that of an engineering model development and not a detailed understanding at the cell level. Frankly, we are not interested in the lower level detail, only gross modeling of cells, genes, and their proteins. They will become the inputs, outputs and control mechanisms of our design approach.

3.4.1 Plant Cells

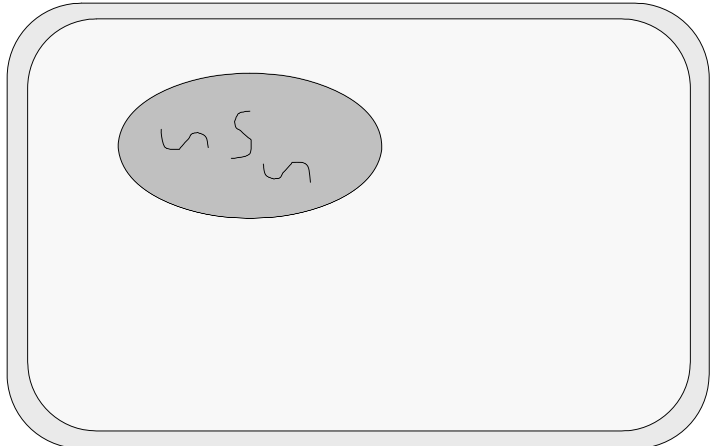
Plant cells are a class of eukaryotic cells which are characterized primarily by have a rigid cell wall. In almost all other ways they are similar to animal cells. Plants generate all of the amino acids they need for protein generation unlike animal cells but other than that,

[16] See Griffiths. This is an excellent overview of genetic analysis.

for our purposes, they function very much the same. Thus as we develop a model for plants the model has no restrictions in its applications to animals as well.

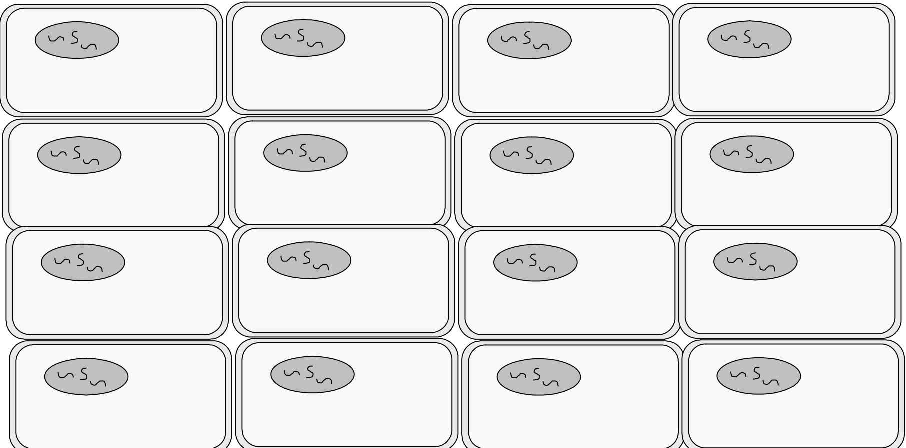
The typical plant cell is shown below. The cell wall and the nucleus are depicted.

Plant Cell



When we look at a collection of plant cells they appear as below. They are aligned and interconnect via various channels. Unlike animal cells plant cells have a much more rigid structure due to the cell wall however the general intercell signaling is identical.

Plant Cell Matrix

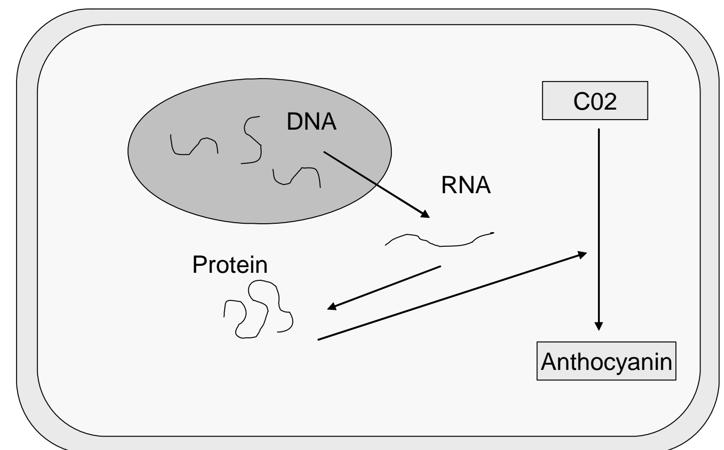


Our interest will be to focus on both the intracell and intercell signaling and control of the pathways.

3.4.2 Plant DNA

Plant DNA processes are almost identical to those of animals. The graphic below summarizes the view we shall take. Each cell has DNA and the DNA uses a mRNA to create proteins. The proteins are then used in the management of the pathways to create the secondary products of the cell, in our case the anthocyanins.

Plant Cell DNA Process



For a single cell the model is quite straight forward. Gene expression causes RNA which causes Protein, which is enzyme in anthocyanin pathway generating the anthocyanin.

We do however want to stress certain issues. There are two extreme views of cells:

Micro/Time View: The micro view looks at a cell at each instant of time and considers what is happening. Is the cell generating a protein and a secondary and if so how and what is the sequence in which this process occurs. It is a focus on a single cell over some time period and we see many things happening.

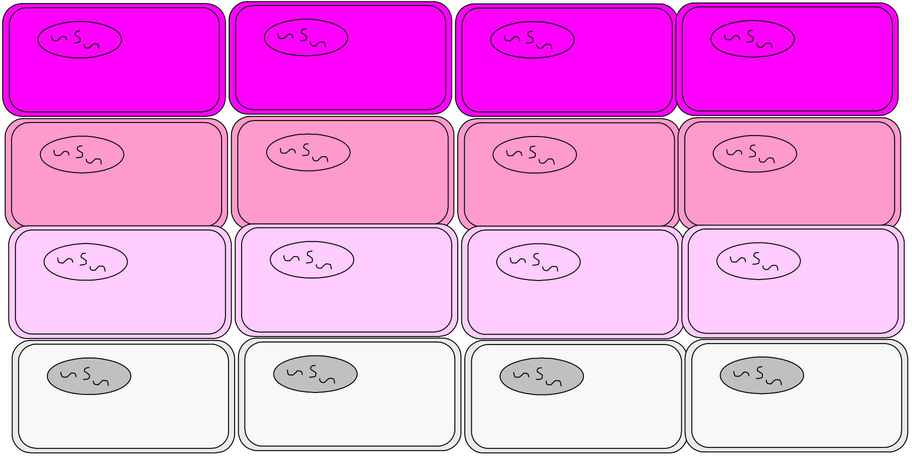
Ensemble View: In this case we look at the cell on average. Namely we say a cell can "on average" produce a protein and can then in turn produce a secondary.

These two views have analogs in mathematical analysis; they are the time averages versus the ensemble average. In mathematical statistics we have the concept of looking at a single cell and time averaging say the concentration of a certain secondary. We know how it is produced and thus over some time window we can look at the average of say pelargonidin and we than measure its average value. In contrast we can take a collection of similar cells and measure the pelargonidin in each cell and take that average. The latter is called the ensemble average. The equivalence of the two is called

the Ergodic Theorem and was developed by Norbert Wiener[17]. The microbiologist typically focuses on the time view. We in this Chapter will focus on the ensemble view. The latter view will allow us to model, predict and control large collections of cells.

Now the figure below depicts a typical problem we want to understand. Consider an array of cells. Consider that they are arranged in ascending order up the petal of the flower, from base to outer edge. Consider now that at each vertical increase that the cells at the same level all have the same color yet at each level they have a differing shade of color. This implies that the anthocyanin concentrations are different at each level but identical at each cell within a level. We will assume we can understand a single cell from our discussions in the last section, if we understand the pathways and their enzyme controls. Now we ask how does one create a mathematical system model which can "explain" the color patterns we see below. This will be a critical question to answer.

Plant Cell Matrix Colors



How do the cells communicate? Why does one cell generate more anthocyanin than other cells. Why is this not just random? What is the control mechanism?

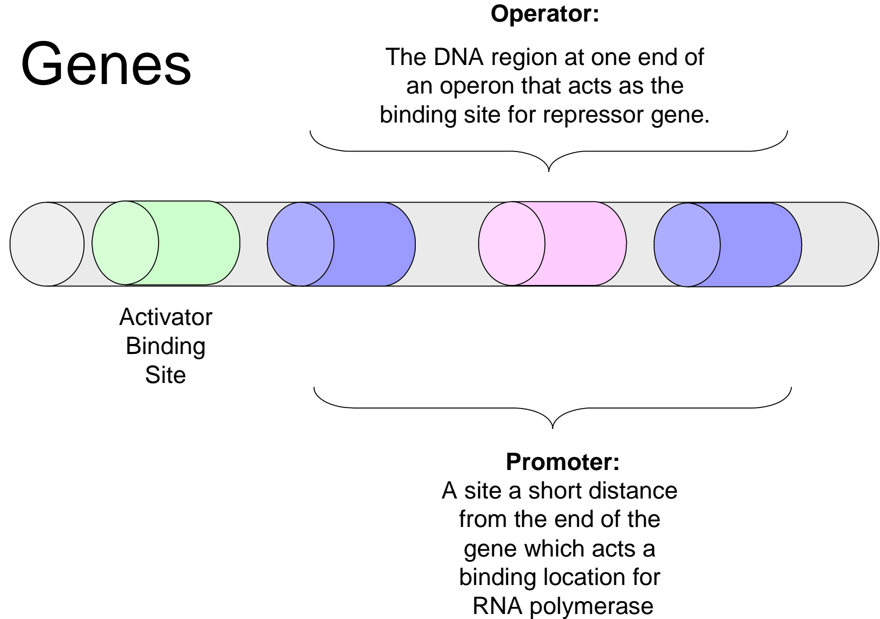
Before we can answer this question we need to delve a bit deeper into the genetics of gene expression.

3.4.3 Plant Gene Processes

The processes in plant genes are generally identical to those in animal and thus human genes. The figure below shows a typical gene structure along with key sites. This

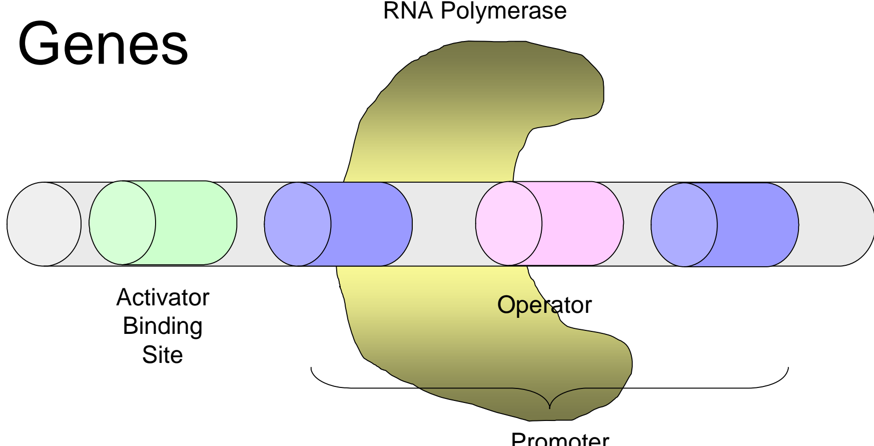
[17] See McGarty, Stochastic Systems and State Estimation.

structure shows the gene activator site which is where activator proteins can bind to start or enhance the expression of the gene. The operator sits and the overall promoter sequence are shown down from the activator site.[18]



Genes express themselves with the assistance of RNA polymerase. The RNA polymerase is key in that it binds to the DNA and then opens it up to allow for the transcription creating the mRNA required for the translation process. In the figure below we show this process.

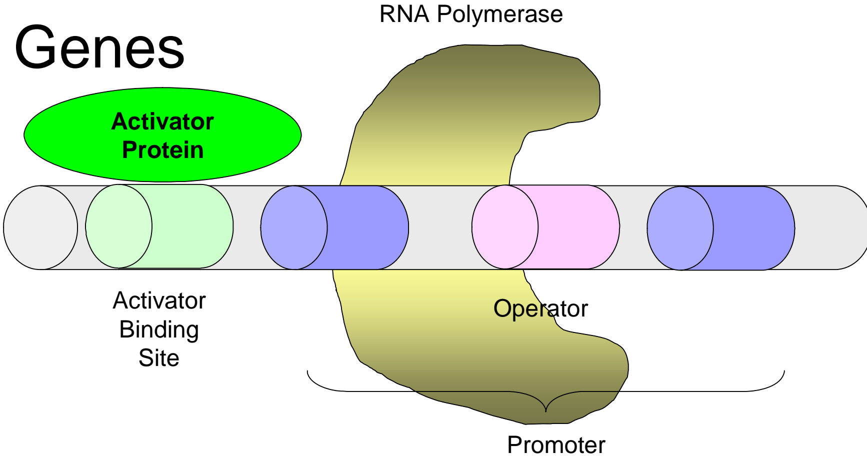
[18] This is detailed in Watson et al. Also see Griffiths et al.



We will now focus on two actions which control the gene expression; activators and suppressors.

3.4.4 ACTIVATORS

Activators are proteins which when attached to the gene assist in the expression of the gene. An activator is a protein resulting from another gene which can assist and facilitate the expression of a gene. Remember we want to look at the ensemble view, not the time view. Thus we assume that the RNA polymerase is continuously acting to produce proteins and that there is a continuous flow at some level of the activators. The cell process from the time view is shown below. An activator binds facilitates the RNA polymerase binding which in turn produces the mRNA and then in turn the proteins via the translation process.



If there is an activator then the gene can be readily expressed. The RNA polymerase then binds, creates the mRNA and this in turn produces the related protein. Activators stimulate this process. The Figure below depicts the location of the gene downstream from the activator and the promoter.

Genes

Genes

Activator Binding Site

Operator

Promoter

Now it is important to understand the activator from a time perspective and then from the ensemble perspective.

1. Activators are proteins generated by other genes in the cell.
2. Activators bind to the DNA and facilitate the production of the gene, which in turn produces another protein.
3. Activators can bind, release and then rebind. Each time they do that they produce another mRNA and that in turns produces another protein molecule.
4. From a time perspective, it is activator, produces gene reading, produces mRNA, and produces protein.
5. From an ensemble perspective we have a concentration of activator proteins and then we get a concentration of result proteins.

This then leads to a simple model:

$$P_o = \text{Output Protein Concentration}$$
$$P_i = \text{Input Protein Concentration}$$
$$P_o = A_{o,i} P_i$$

But there is also a dynamic model which we can state; to some degree this model is a hybrid of the time and ensemble approach. The model states:

$$\frac{dP_o}{dt} = f\left(P_o(t), P_i(t), t\right)$$
$$P_o(0) = P_o^0$$
$$P_i(0) = P_i^0$$

Now we must remember that this simple two protein, two gene model is just a simplification. In reality we may have dozens of not hundreds of genes in this process. Now consider a simple linear model for this two gene system:

$$P_i(t) = P_i^0 \exp(-\lambda_i t)$$

$$\frac{dP_o(t)}{dt} = A_{o,i} P_i(t) + A_{o,o} P_o(t)$$

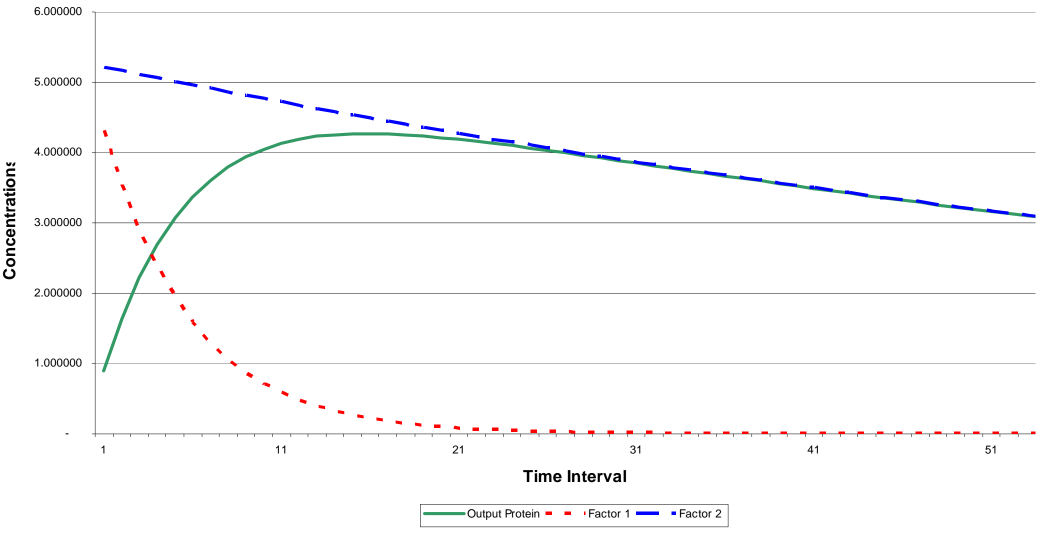
We can solve this differential equation. It is:

$$P_o(t) = k_{o,i} P_i(0) \left[\frac{\exp(-\lambda_i t) - \exp(-k_{o,o} t)}{\lambda_i - k_{o,o}} \right]$$

$$where;$$

$$A_{o,o} = -k_{o,o}$$
$$A_{o,i} = +k_{o,i}$$

We have solved this for a simple example using constants of 0.01 and 0.2 respectively.

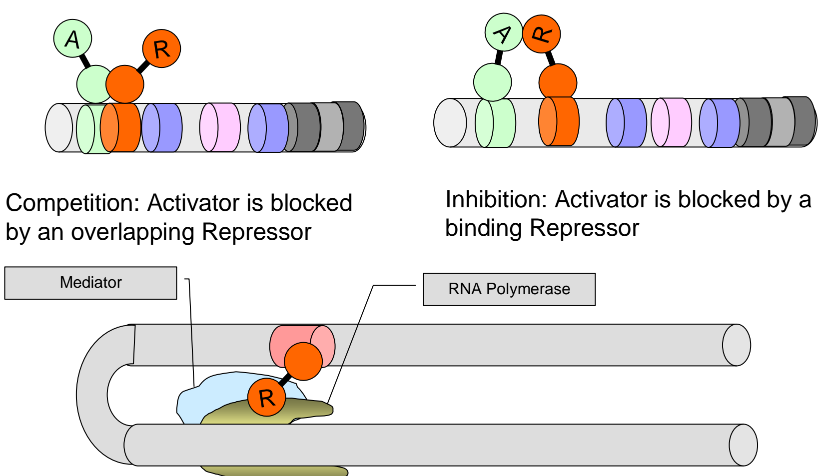


Note that the output protein concentration reaches a peak and then decays as per the driving protein. We will see this phenomenon again.

3.4.5 Repressors

In contrast to activators we also have genes which are suppressors. Three methods of suppressor action are shown below. A suppressor does the opposite of an activator. It suppresses the expression of a gene. The same logic will follow the repressor as was with activators. We again also want to view this from an ensemble perspective.

Repressors



Direct Repression: Repressor blocks transcription

As we did with the activator, we see a repressor stops the generation of the protein. This it is nothing more than a negative driver to protein generation.

3.4.6 Summary of Actions

We can now summarize what we have presented here:

1. Color is the result of anthocyanin production.
2. Anthocyanin production is a product of a specific pathway.
3. Pathways are mediated by enzymes, which are proteins generated by genes in the cell.
4. Proteins are generated by genes.
5. Gene activation is modulated by activator proteins and suppressor proteins.
6. Activator and suppressor proteins are generated by other genes.
7. One can model this overall process by a linked set of equations, both of a time varying nature and an ensemble, average steady state, nature.
8. An overall state model can be developed for the genetic control of color in plants.

We can now take this set of conclusions and use it to construct the state model.

3.5 Expression Analysis and Implications

In this section we develop a systems approach to the problem of color analysis and synthesis. This work is based upon the recent work of Szallasi and others. However this also builds upon the work in McGarty (1971) which focused a systems approach to the overall identification problem.

3.5.1 Approach: Engineering versus Science

The approach we take in this Chapter is an engineering approach rather than a biological approach.[19] Our interest is in developing a model or sets of models which allow us by a verifiable means to show how the genes react and interact to produce the plant colors. We can compare this to the engineering approach to circuit design of transistor circuits versus the science of understanding the semiconductor from the point of view of detailed quantum mechanical models.

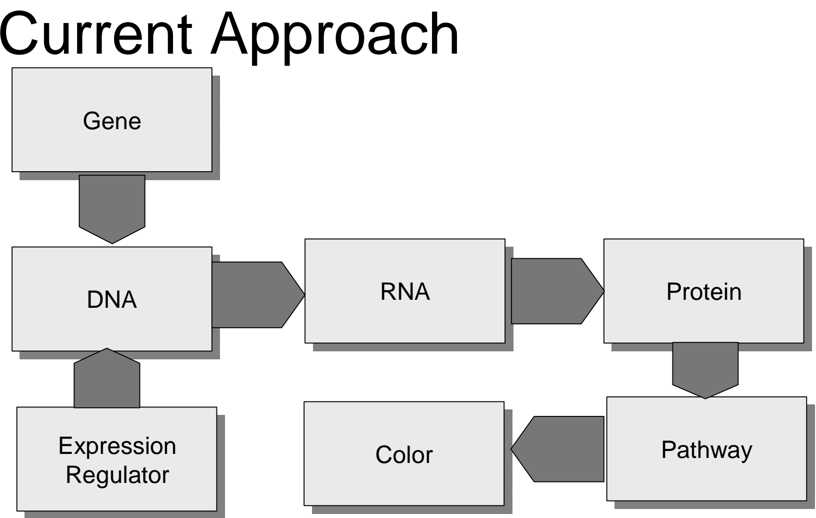
The biologist in our approach is akin to the physicists and engineers who approach the cell from the bottom up, trying to understand all of the intricate processes and steps that lead at the micro level to the developments we look at herein. In our approach it is akin to the engineer knowing that there is some function inside the semiconductors which may clearly be important but the engineer's interest is in designing and analyzing the transistor as a circuit element.

Thus for an engineer, if we increase a current here we get a decrease or an increase at some other point. The engineer creates a world view of a macro set of processes and models the details of the biologists in our case with a few set of equations which show the results of increases and decreases. This model must then be valid table and verifiable. One must be able to make measurements to show that the processes predicted indeed occur, to a reasonable degree of accuracy. Then one can analyze a genetic circuit and then in addition one can design a genetic circuit. We then can understand where the colors come from and possibly engineer the genes to develop and deliver on colors we desire.

3.5.2 A Control Paradigm

[19] There has been a significant set of development recently in analyzing genetic data from a systems perspective. In this Chapter we have taken such an approach. The recent work by such authors as Perkins et al, Vohradsky, Hatzimanikatis et al, and the recent book by Szallasi are seminal. However, there is an issue here also or world view and what does one really want from the analysis. The bench scientist looks to understand all the details of the underlying processes. The engineer seeks to understand enough to model the process and to do so with a reasonable degree of accuracy but the ultimate goal for the engineer is control of the process and generation of new processes.

The basic control paradigm is contained in the following Figure. The expression regulator may be an activator or suppressor. It may be a result of a gene expression in the cell itself or quite possibly as we shall discuss fed through from another cell. There are many of these regulatory cycles and they are all interconnected. This basic paradigm is one of hundreds or thousands of such interconnected flows.

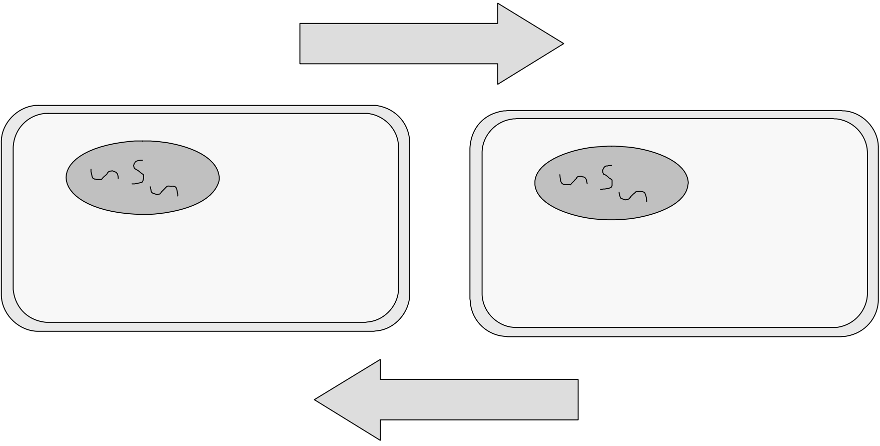


In developing our models we will use this construct. However, we can frequently focus on natural clusters of related genes. They may be a dozen or more such related genes in each cluster and possibly hundred of such clusters. Although cells and their proteins may affect all other cells, only a few of the genes regulated have a significant level of regulation. The low levels of "regulation" we shall consider just as noise.

3.5.3 Cell Signaling: Intra and Inter Cell

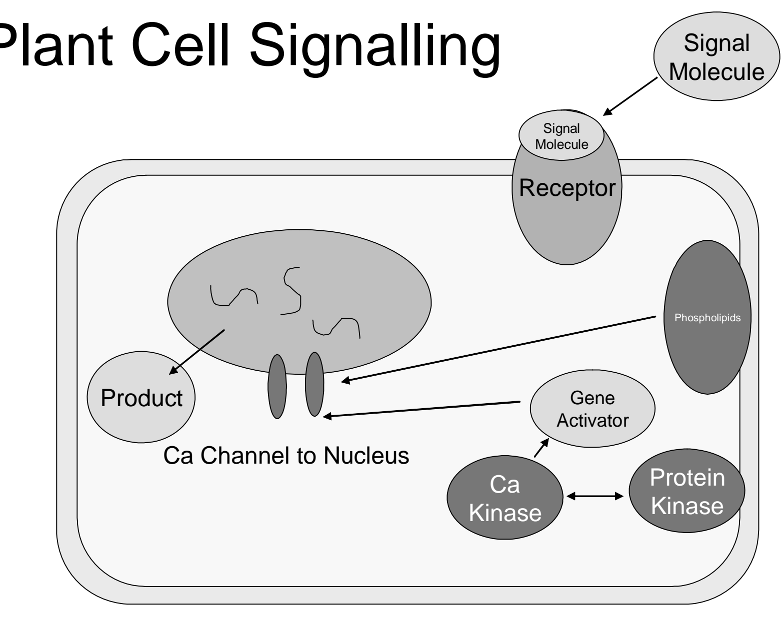
We must also better understand the inter cell signaling. Although we include it in this Chapter we have not as of yet produced a robust enough model for this set of processes. The Figure below presents the essence of the problem.

Plant Inter Cell Communications



What do the cells use to communicate and how. What are the elements?
Proteins?

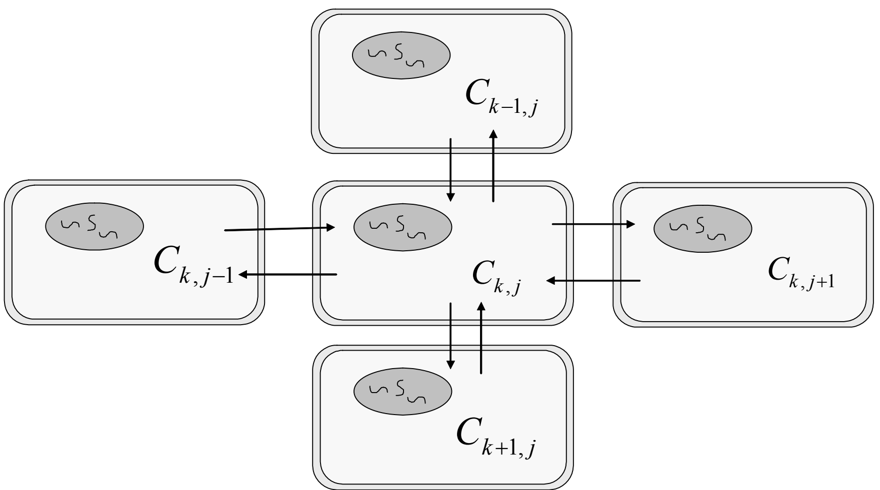
Key to intercell signaling will be the receptor elements which control the flow of the controlling elements. This means that we must be able to introduce certain additional elements in the model which at this time are not yet fully developed. The Figure below highlights the issues of concern in this area.



Dey, Plant Biochemistry, p. 373, 1997 Academic

Then we must be able to establish a full network view of the signally processes. There has been considerable work looking at this from a meta perspective as some neural network. However the approach does not yet provide an adequate refection of a gene by gene analysis.

Plant Inter Cell Communications



What do the cells use to communicate and how. What are the elements? Proteins?

3.6 Flower Color Expression

We have just shown that there are a wide variety of coloration in the daylily. In a little over a hundred years we have taken the dozen or so species and intermixed them and as a result have created a very complex set of flowers with characteristics which differ dramatically from the species.8F[20] The species have managed to maintain their separate identities over thousands of years but in a small fraction of time we have been able to introduce multiple forms and colors. To understand this process we first have to understand where the colors come from. How do we get purple from a plant which is red, yellow, orange and possibly even brown? How are the colors made and how do we get from there to where we are today.

---

[20] See Lensaw and Ghabrial for an excellent discussion of the tulip. In contrast to the daylily, the tulip craze of the seventeenth century was a dramatic bubble, and the irony was that most of the color variations were induced by viruses.

The first step in understanding that process is to understand the pathways that lead to color production in a single cell. Then we can address the issue of multiple cells and finally how the cells communicate. How do we get an eyezone for example. Why if a cell is whit do we go so abruptly to a purple eyezone. What is the mechanism for this process? We begin the exploration of this issue with a analysis of the underlying pathways.

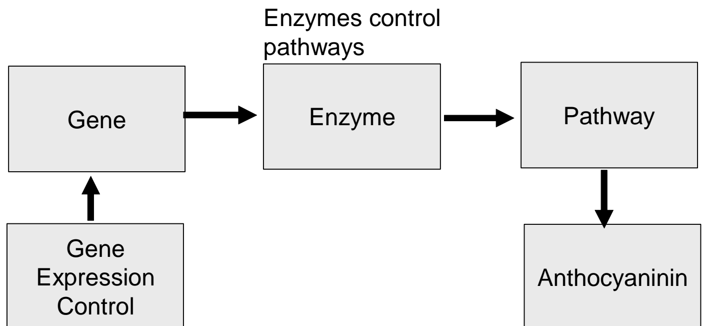
3.6.1 Pathways and Enzymes

Pathways are nothing more than a set of chemical reactions which get us from some primitive chemical to a more complex but useful chemical structure.9F[21] In fact the pathways may be just a set of processes going from any one chemical structure to another independent of the nature of the starting and starting chemical. Some pathways are linear going from a beginning to an end and some are circular taking us from the beginning and back again; the Krebs cycle is an example. What makes the pathway work? Just three elements are required: (i) the underlying chemical constituents, (ii) some form of energy, (iii) generally some form of facilitation such a catalyst and in our analyses this is an enzyme.

The general flow structure we look at is shown below. In our view, not the only such view but one convenient for the development of our argument, we have the pathway but it facilitated by an enzyme, a protein. The protein is generated by a gene. And the gene is activated by some other element, generally another protein. In our case shown below the output is some anthocyanin. The more of the enzyme, namely the more the gene expresses itself the more anthocyanin we get. Thus if we can get the gene to express then we get more of that specific anthocyanin, more pelargonidin for example. We defer to the next section how we get this gene to express so strongly.

[21] See Taiz for an excellent overview. Dey is also a superb and current reference. The older references by Goodwin are useful but they fail to account for the genetic effects.

Pathways, Enzymes and Expression



Many factors control the expression of the gene. Even the cell which is next to the one producing the enzyme.

Each anthocyanin creates a color element. The more of that one type the richer that element. Combining them together creates a totally new color.

The opposite is also true. Namely if we can suppress the gene then we can get less and even possibly no anthocyanin from the pathway. This is the first step in the development of an overall system model.

3.6.2 Anthocyanins

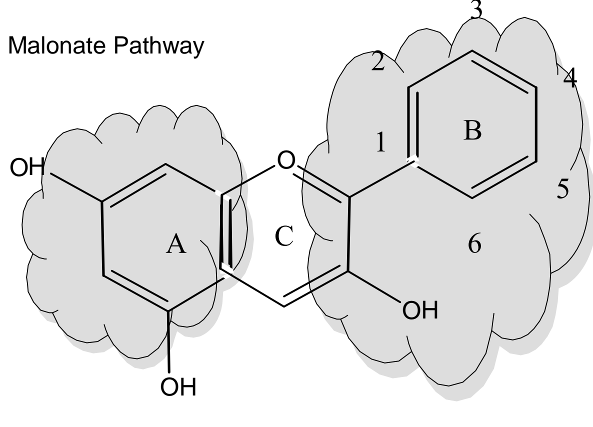
Let us consider our first pathway. This is the pathway which creates anthocyanins.10F[22] The anthocyanin molecules is shown below. Note on the B ring we have six sites to which we can attach differing molecular chains. This will be an important element when we see the different configurations and their implications.

[22] See the Chapters by Mol and also by Winkel-Shirley. They are excellent in the characterization of the pathways. Also the Chapters by Holton and the one by Jaakola are quite useful here as well.

Anthocyanidin

The anthocyanin or anthocyanidin molecules comes from two different pathways. In the figure below we have taken the basic resulting molecule and have shown that there are two elements; one is from the shikimic pathway and the other from the malonate pathway. This means that we have to understand both pathways to understand the ultimate abundance of the product.

Anthocyanidin



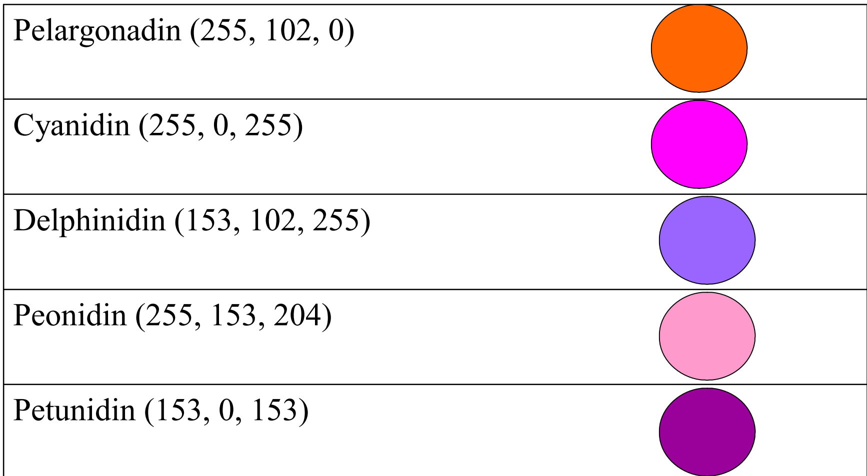
Before continuing we want to look at what the results would look like if we have different substitutes on the B ring. In the Table below we show that the terminations on the 3, 4 or 5 elements yield different results. The results give pelargonadin, cyanidin, delphinidin, peonidin, and petunidin. Each obviously named after their related flower and each resulting an anthocyanin of a different color.

Colors

Anthocyanidin	*Substituents*	*Color*
Pelargonadin	4'-OH	orange-red
Cyanidin	3'-OH, 4'-OH	purplish red
Delphinidin	3'-OH, 4'-OH, 5'-OH	bluish purple
Peonidin	3'-OCH_3, 4'-OH	rosy red
Petunidin	3'-OCH_3, 4'-OH, 5'-OCH_3	purple

In the Table below we have shown the colors of each of these as well as the weighting of a red, green and blue combination which best matches the color. Thus one can in an 8 bit color schemes, as one would find in any PC color scheme, get the resulting anthocyanin colors by blending the R, B, G elements to yield what we are seeking. This relating the colors back to RBG is critical since it get reflected in the ultimate flower color.

Colors (R, G, B)



Pelargonadin (255, 102, 0)	
Cyanidin (255, 0, 255)	
Delphinidin (153, 102, 255)	
Peonidin (255, 153, 204)	
Petunidin (153, 0, 153)	

Now if we assume we have only anthocyanins for color, and that we have the above combinations available, we ask how do we combine these colors in a weighted manner to obtain the desired color. This approach is critical to the overall understanding. First we show by a weighted RBG we get the color we seek or the color which is presented. The we assume that if we can then do the same for each anthocyanin, then we can create any desired color from a weighted collection of anthocyanins. This means that we can then determine what the relative percents of expression of any anthocyanin is and this lets us then go back to how strongly the gene for that anthocyanin is expressed. The model we presented earlier will be a key element in this overall process.

3.6.3 Other Color Elements

Anthocyanins are not the only elements which are secondary products which produce color. There are three classes of chemicals which give rise to color; anthocyanins, flavones or flavonols, and carotenoids. The Table below depicts the different elements and their colors. The approach we took above for the anthocyanins can be take for the flavones and carotenoids as well. It should be noted that there may not be a unique solution here but there are several possible but they can be narrowed down by actual determination of one to three elements as baseline.

Class	*Agent*	*Color*11F[23]
Anthocyanidin		
	Pelargonidin	orange-red
	Cyanidin	purplish-red
	Delphinidin	bluish-purple
	Peonidin	rosy red
	Petunidin	purple
	Malvinidin	
Flavonol		
	Kaempferol	ivory cream
	Quercetin	cream
	Myricetin	cream
	Isorhamnetin	
	Larycitrin	
	Syringetin	
	Luteolin	yellowish
	Agipenin	Cream
Carotenoids		
	Carotene	orange
	Lycopene	Orange-red

We now summarize the other element classes.

3.6.4 CAROTENOIDS

Carotenoids are what is quite common in the carrot, the orange hew we see in that root. Its molecular structure is shown below, this is beta carotene.

[23] See Taiz p. 334 for the anthocyanidin color and Bernhardt for the flavonol and carotene.

3.6.5 Flavones

The flavonols, or flavones are quite similar to anthocyanin. Their structure is shown below. Note that we have compared it to that of anthocyanin.

Anthocyanidin

Flavonol

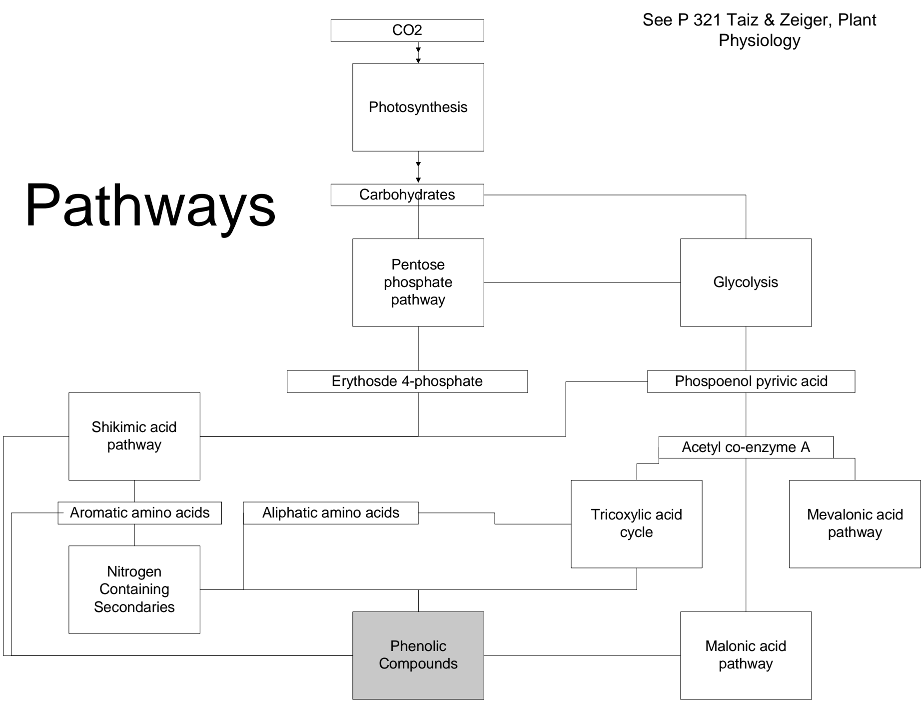
We can also show how closely they relate in substitutions and colors. This is shown in the Table below.

Flavonol	Anthocyanidin	Substitution	
		3’	5’
Kaempferol	Pelargonidin	H	H
Quercetin	Cyanidin	OH	H
Myricetin	Delphinidin	OH	OH
Isorhamnetin	Peonidin	OCH_3	H
Larycitrin	Petunidin	OCH_3	OH
Syringetin	Malvinidin	OCH_3	OCH_3

3.6.6 PATHWAYS

In this section we present the pathways for the three classes we have described above. We first present an overview of the pathway and then we present the details of the pathway and the enzymes used in each step. The key observation is that we must have enzymes to go from step to step in the pathways and that if any one enzyme is missing we cannot proceed on that path, and further the path with the small amount of enzyme becomes the limiting path. Thus, we do not have a one to one map here. The production of any one anthocyanin, for example, if limited by the lowest produced enzyme, and the other enzymes may be present in abundance.

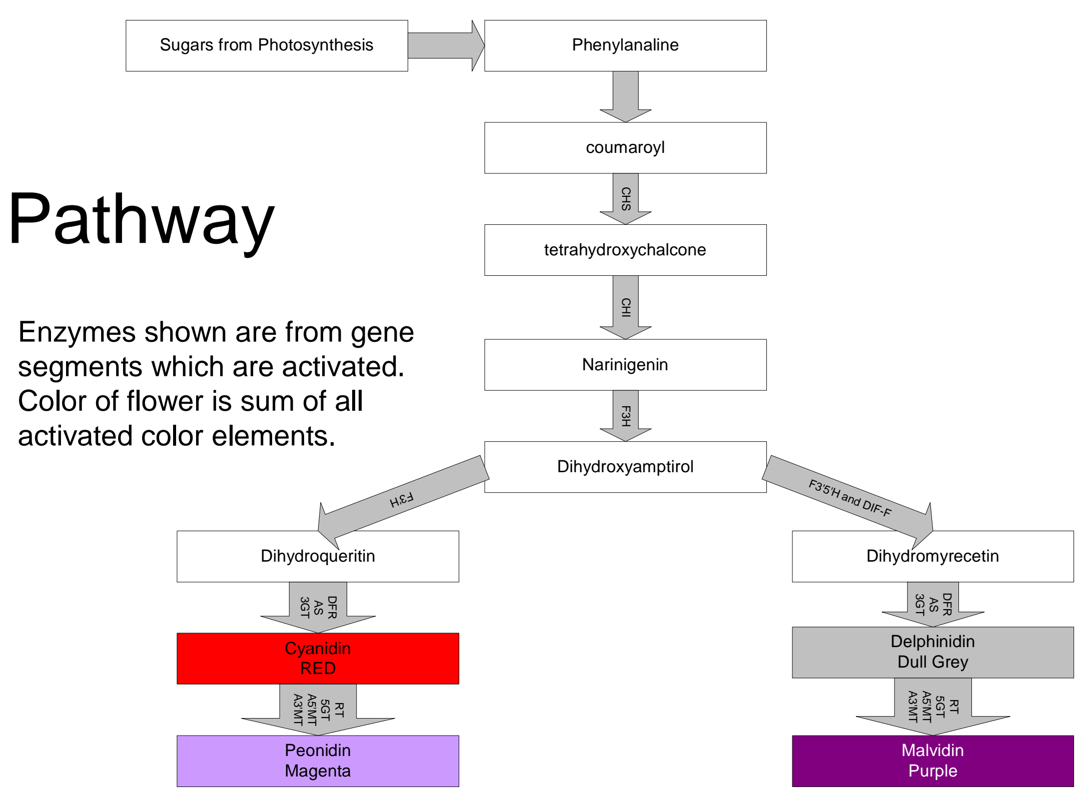
The following is the overall pathway for all elements.



The above shows how we start from CO2 and then go through a variety of other pathways. We will review those pathways in some detail since it is the enzyme control in them which is key.

3.6.6.1 Anthocyanin Pathway

The anthocyanin pathway with the controlling enzymes is shown below. The enzymes are presented in the arrows linking each step in this pathway. This pathways shows the start as a sugar element and then goes to phenyanaline and then down through the chain to one of the four indicated anthocyanins.



Note that at each step there is an enzyme element. The genetic loci for cloned flavonoid enzymes in Arabidopsis are shown in the following Table.12F[24]

[24] See Similar information for maize, petunia, and snapdragon is described by Holton and Cornish (1995). Based on the AGI map, 11/12/00; numbers in parentheses refer to P1 or bacterial artificial chromosome clones on which these sequences reside. Transposon- tagged mutant for FLS1 (Wisman et al., 1998).

Enzyme	Locus	Chromosome	Map Position
CHS	tt4	5	7,050 kb (MAC12)
CHI	tt5	3	21,000 kb (T15C9)
F3H	tt6	3	19,600 kb (F24M12)
F39H	tt7	5	4,400 kb (F13G24)
FLS	fls1<Enc	5	FLS1: 4,700 kb (MAH20) FLS2-5;: 32,150 kb (MBK5) FLS6: 24,350 kb (MRH10)
DFR	tt3	5	23,800 kb (MJB21)
LDOX	tt19	4	16,900 kb (F7H19)
LCR	ban,ast d	1	26,800 kb (T13M11)

The pathway for the conversion of the sugar erythrose to penylanaline is shown in the reaction below. This accounts for the upper part of the pathway which we have shown. It uses the Shikimic pathways which we have shown in the initial discussion on the pathways.

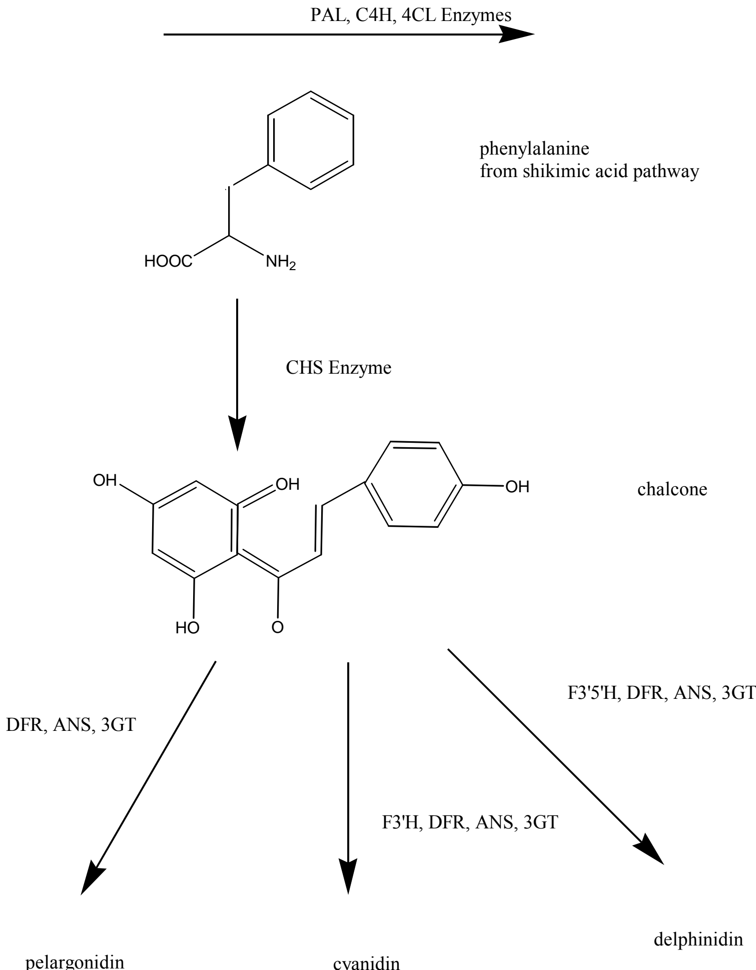
D Erythrose 4 phosphate

Shikimic Acid

Phenylalanine

Tyrosine

The conversion details from pehnyanaline through chalcone to the anthocyanins is shown in the reaction below. We have reiterated by transition the enzymes which facilitate each step in this process.



What these process point out can be summarized as follows:

1. There are common pathways which are operational in all plants for the generation of the pigments.

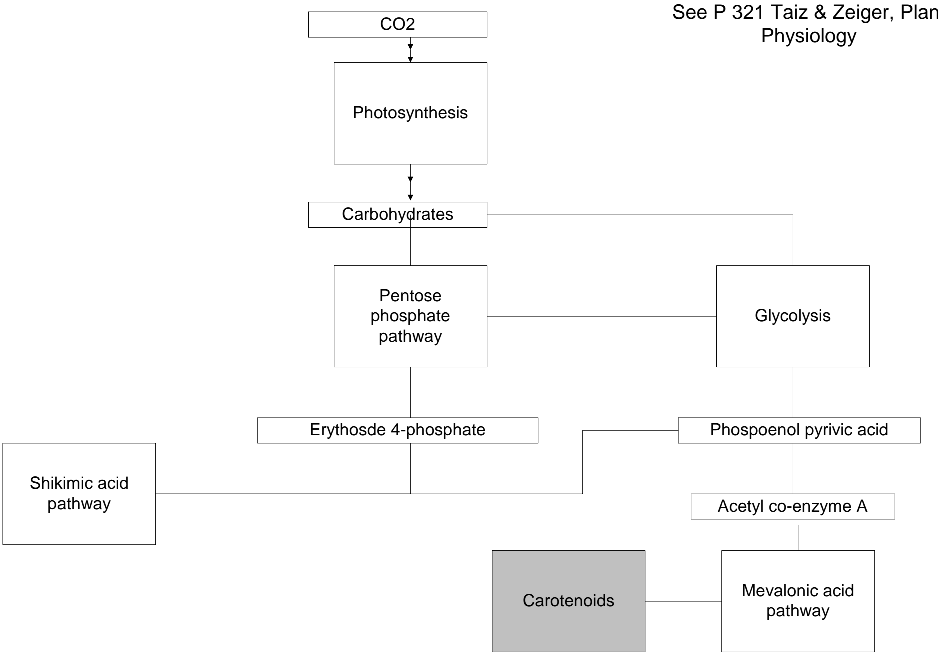
2. Enzymes used as activators modulate the amount of production of the enzymes.

3. The products of these pathways, the anthocyanins, are driven by the concentration of the facilitating enzymes.

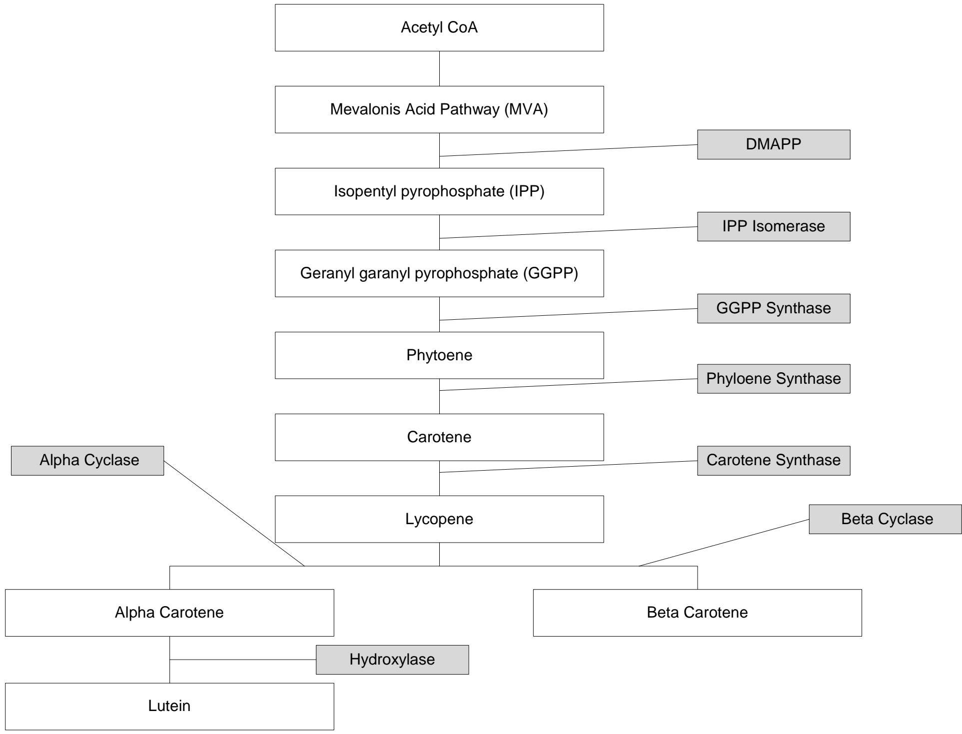
Secondary products always have this type of production process. As we look at a cell, from a system point of view we see facilitating proteins and secondary products. The concentration of the secondaries are proportional, in some general way, to the concentration of the facilitating proteins. However we see there are many facilitating proteins which may make this a more complex analysis, however doable.

3.6.6.2 Carotenoid Pathway

We have shown the carotenoids as above. The carotenoid pathway is shown below. We have demonstrated this in general terms earlier but in this case below we see the specific details.



We show below the pathways and the facilitating enzymes. In many ways it appears identical to the anthocyanin pathway and the facilitating enzymes.



3.6.6.3 Flavonol Pathway

The flavonol pathway is identical to that of the anthocyanin. See the work of Winkel-Shirley.

3.7 Conclusions

The basic principles of genetics is essential to understanding the overall issues of species and their related colors, patterns forms and growth. The principles that we believe are critical are those related to activators and repressors and the interlinking of them in a dynamic system. As we have stated herein, we look at this as an engineering and design problem and not as a scientific assault on the unknown. As engineers we look and say; what can we achieve with what we know, and if that does not work then go back and question the assumptions. The alternative is to make what we don't know acts as "noise". That is the engineering approach.

The issues of the secondary pathways will become essential when we look at the issues of how genes control, the pathways and then how the pathways generate the color we see in flowers. We have expanded upon this extensively in two areas; color and its control and patterns and their control.

4 PHYLOGENETICS, DNA, CLASSIFICATION

The genus Hemerocallis has had various attempts at classification since the time of Stout. The primary approach has been via phenotypic methods and Erhart has recently proposed an alternative classification to Stout. With the introduction of various genetic methods for classifying the genus, a dendrogram has been proposed using the AFLP methods of genetic sampling. This paper reviews the various techniques and then also reviews and summarizes the several phylogenetic approaches which have been taken to the present. The paper then details some analyses of comparisons of the papers in the literature and makes suggestions as to potential next steps. To data there has not been a detailed genetic analysis which has allowed for detailed evolutionary classification.

4.1 INTRODUCTION

The classification of plants involves many complex methodologies. Recently, with the use of DNA methodologies, there has been a re-assessment of many of the classifications based upon morphology and the other more classic limited metrics and measurements. This paper focuses on classifications using those newer techniques. It looks at what the methodologies are and evaluates their respective advantages and disadvantages.

It then looks at the methodologies of taking the data collected from the DNA methodologies and then creating classifications based upon that data. Certain methods use generally gross level methods and others use quite detailed and sophisticated methodologies which attempt to incorporate actual DNA modification. The eventual goal of this paper is the set down a classification of the genus Hemerocallis which reflects the best current thinking using DNA measurements.

We begin the discussion with the posing of several key questions and then we try to answer them with the tools currently available. There are many such questions which can be easily posed but not readily answered. The genus Hemerocallis is a simple genus of a monocot plant which is originally from Asia, including China, Japan and Korea. It is a common plant which comes in a variety of species.

4.2 KEY QUESTIONS

There are many questions which beg the answering. The following are several:

1. What are the species in the genus Hemerocallis? In fact, what do we mean by species?

Ernst Mayr was famous for defining a species as a collection of living organisms which have the capability of interbreeding. (See the various works of Mayr on the issue of

species) Elephants and lions do not interbreed, thus it is obvious that they are different species. The genus Pinus and the genus Picea do not interbreed, thus they are composed of different species. Yet the species of Hemerocallis readily interbreed, begging the question of species as posed by Mayr. Hemerocallis all have 11 pairs of chromosomes, namely they are diploid with a total of 22 chromosomes, with the exception H fulva whish is a diploid with a total of 33 chromosomes.

2. Within what one may see as a species, how much variation can one tolerate and still call it a species?

This is a key question. Is there a specific characteristic which defines a species? For example, H dumortierii has brownish sepals, and it is sessile. If the sepals are no longer brown is this now a new species? Or is it just a variation? What is controlling the color, is it a definite species characteristic?

3. What characteristics do we look for to distinguish a species one from the other? What are the most telling of the characteristics, and why does one select those characteristics?

When we look at H fulva we know it has 33 chromosomes. That makes H fulva unique. It also is generally sterile. Then we have species which all bloom at about the same time. The H minor, H dumortieri H middendorfii and H flava all bloom at the same time. Perhaps this means that they could inbreed. But dumortierii and middendorfii have sessile flowers whereas minor and flava have branched flowers. Is sessile and branching a major factor which makes them species? H minor is grass like, with drooping scapes and leaves which droop and are thin and short. H flava is erect with larger leaves. Both naturally pollinate by bees. However flava blooms at night and minor is more of a day bloomer.

4. If we can identify a species, and we can see the collection of all species, how can we relate one species to another? Is there some closeness of one species with another, and moreover is there a way to relate them so as to see how they evolved to where they are now? Finally, can we "look back in time" to understand what the ancestor was or the ancestors were?

This is the process of developing a tree showing the relationship of one species to another. The issues of defining the relationship are driven by a closeness measure. It can also be driven by a change in genes. For example if one species has a gene given by:

…CCTTAGCCT…

And the other species has a gene:

…CCATAGCCA…

Then we may ask what ways did these genes get to this point? If we know that genes mutate at the rate of α per thousand years then we could calculate the most likely ancestor of these two genes. This is one of the many ways one can approach this problem. At the heart of any such approach is some measure of closeness. How close are the two genes, how close are two proteins, and so on.

5. Using genetic tools, how would we best approach the issues of identifying species? What are the best genetic markers, and how detailed should one get to optimize the task? Given the best possible genetic marker, how do we then sort and arrange the measurements to assist in defining species?

There are thousands of genes. Which ones should we focus on and should we weight them differently and if so how differently? One can assume that they have the best set of genes from all the species. Then one must look at both intra species matches and interspecies matches.

This paper examines many of these questions. There are answers for some, work in progress for others, and many which are still a way from being addressed.

4.3 Prior Efforts

In the past ten years there have been many studies regarding the genus Hemerocallis. We briefly review a few.

Chung and Noguchi in 1998 published a paper on H middendorffii where they looked at the differences in morphological characteristics over regions in Japan and Korea. This paper provides a good benchmark for the use of morphology. It shows that there is also some significant variation within a species as to the morphological characteristics.

Chung in 2000 collected H hakuunensis samples and using an enzyme technique examined the spatial variability within the species. Three specific enzymes were analyzed and there was significant spatial variation was found in one and little in two others. The species has some variability but not a great deal.

Hasegawa et al in 2006 reported on hybridization between H fulva and H citrina. The fulva is a day blooming plant and citrina a night bloomer. There is some crossing that result from the slight overlap of bloom. Specifically the authors' state: "*Most F1 hybrids showed diurnal flowering. These findings indicate that only a few genes have strong phenotypic effect on the determination of lowering time in Hemerocallis, and suggest that the evolution from a H. fulva-like ancestor to H. citrina was not a continuous process by accumulation of minute mutations*." This

study has been flowed up by Yasumoto and Yahara in 2008 where they deliberately set F1 crosses. The belief is that H fulva is an ancestor to H citrina.

The work by Kang and Chung in 1997 examined the genetic variation in H hakuunensis. The authors used enzyme markers and they observed:

"Hemerocallis hakuunensis, a Korean endemic species, maintains considerably higher levels of allozyme variation within populations …and substantially lower levels of allozyme divergence among populations ….. than average values reported for other insect-pollinated, outcrossing herbs. Indirect estimates of the number of migrants per generation … indicate that gene flow has been extensive in H. hakuunensis. This is somewhat surprising when we consider the fact that no specialized seed dispersal mechanism is known, flowers are visited by bees, and the present-day populations of the species are discontinuous and isolated. Results of a spatial autocorrelation analysis based on mean allele frequencies of 19 populations reveal that only 13% … of Moran's I values for the ten interpopulational distance classes are significantly different from the expected values and no distinct trend with respect to the distance classes is detected. Although it is unclear how the populations are genetically homogenous, it is highly probable that H. hakuunensis might have a history of relatively large, continuous populations that had more chance for gene movement among adjacent populations after the last Ice Age. In addition, occasional hybridization with H. thunbergii in areas of sympatry in the central and southwestern Korean Peninsula may be one factor contributing the present-day high allozyme variation observed in H. hakuunensis."

The Kang and Chung study is one of the first to detail genetic markers.

Kang and Chung in 2000 looked at the high levels of enzyme variation within a population and low divergence within and amongst species. This was done for H. thunbergii, hakuunensis, hongdoensis, taeanensis, middendorffii, thunbergii. Specifically the authors' state:

"Hemerocallis thunbergii, H. hakuunensis, H. middendorffii, and H. taeanensis had high genetic diversity. On the other hand, three populations of H. hongdoensis maintained significantly … lower mean values of HEe…. than those for the other four Hemerocallis species. Hemerocallis hongdoensis also had the lowest number of alleles…"

"As expected, Korean populations of H. thunbergii and H. middendorffii have high genetic diversity. The two species have a wide geographic range distributed from China to parts of the Korean Peninsula and the Japanese Archipelagos. Most Korean populations of H. thunbergii grow commonly in the open, grasslands on hillsides in the southwestern Korean Peninsula. It has been

observed that Korean populations of the species are large and have a relatively continuous distribution."

From the Kang and Chung paper they provide a classification based upon the enzyme studies as follows:

Fig. 3. UPGMA cluster analysis of 30 populations of *Hemerocallis* species in Korea based on Nei's (1972) measure of genetic distance. Abbreviations are from Fig. 1 and Table 3. *Hhak, H. hakuunensis*; *Hmid, H. middendorffii*; *Hthu, H. thunbergii*; *Htae, H. taeanensis*; and *Hhon, H. hongdoensis*.

In the above classification, what are grouped are the intraspecies and the interspecies. The grouping methodology, UPGMA, is discussed herein. The classification demonstrates several key facts:

1. Intraspecies variation can be significant. In the above we have five species and we can see that the H hakunensis has substantial intra species variability.

2. Interspecies variation is also quite extensive. It is not at all clear from this dendrogram how far back in evolutionary time the species split but we can see that H middendorfii and H hakunensis are related as are H thunbergii and H taeanensis, whereas H hongdoensis is not. The question then is which is closest to the true ancestor.

Tompkins et al in 2001 published the first paper on the use of AFLPs to determine the genetic variation in Hemerocallis. We will focus on their work latter in the paper. Their study presents one of the first truly comprehensive genetic dendrograms or classifications of the genus.

Guerro et al in 1998 performed a detailed genetic analysis of the specific genes which controlled senescence. They used cDNA genes for this specific purpose. This appears to be one of the first truly gene studies and one of the first to create cDNA for the genus.

4.4 The Problem of Classification

Classification of species has been at the heart of all plant systematics. The classification process generally tries to arrange plants into a logical form and doing so to sort the species in some evolutionary manner. Thus the magnolia is a more distant entry into the angiosperms and the asters are more recent. This conclusion is based upon the appearance of certain morphological characteristics found in what may now be extinct plants. One may see a certain characteristics in a magnolia which is found at period X and then see the aster characteristic in period Y and Y is more recent than X and thus the asters are in an evolutionary sense a more recent group then the magnolias. This is a simplistic way to explain the process.

This type of classification works well on families and possibly on genera, if at all. It seems not to work well on species because the historical evolutionary evidence is lacking. Thus species are related purely by the current morphological characteristics.

In the genus Hemerocallis one common characteristic could be sessile flowers versus branched flowers, an approach taken by Stout. At the other extreme would be the analysis of the genes of various species and then to attempt to relate one to the other.

The issue of genetic relating can be phrased as follows:

1. The genus Hemerocallis has 11 chromosomes and 22 chromosome pairs in the diploid species. There is estimated to be several thousands of genes, and the gene length may vary from dozens to hundreds of nucleotides.

2. Certain of the genes have been identified and certain of them are common across other families, such as the genes controlling the secondary pathways of the pigment sources.

3. If we were to look at a large enough collection of genes, and then compare them both within and between species it would be possible to characterize the species based upon the genetic consistency. Current methods in bioinformatics would allow for the assessment of consistency across the gene structures.

4. Using tools that have been developed which account for the changing of genes due to various mechanism one may be able to take the set of existing species and then work backward to attempt to determine how the speciation occurred genetically and how long such speciation may have take and also to determine if there was one or several ancestors. This can be accomplished using the maximum likelihood approach which we discuss herein. Such an approach is highly complex.

Thus it is possible that in time the genetic speciation of Hemerocallis can be elucidated. The state of knowledge at the current time however does not permit that. There has been a great deal of work using other methods which we discuss herein.

One of the important issues to address when performing a phylogenetic assessment is to clearly delineate between intra-species and inter-species variations. In the dendrogram shown below from the work of Kang and Chung (1997) the authors genetically analyzed the species H. hakuunensis and from that analysis demonstrated significant genetic variation within the species. The authors' state:

"Hemerocallis hakuunensis, a Korean endemic species, maintains considerably higher levels of allozyme variation within populations… and substantially lower levels of allozyme divergence among populations… than average values reported for other insect-pollinated, outcrossing herbs. Indirect estimates of the number of migrants per generation … indicate that gene flow has been extensive in H. hakuunensis. This is somewhat surprising when we consider the fact that no specialized seed dispersal mechanism is known, flowers are visited by bees, and the present-day populations of the species are discontinuous and isolated."

The authors used enzyme analysis to develop the following tree.

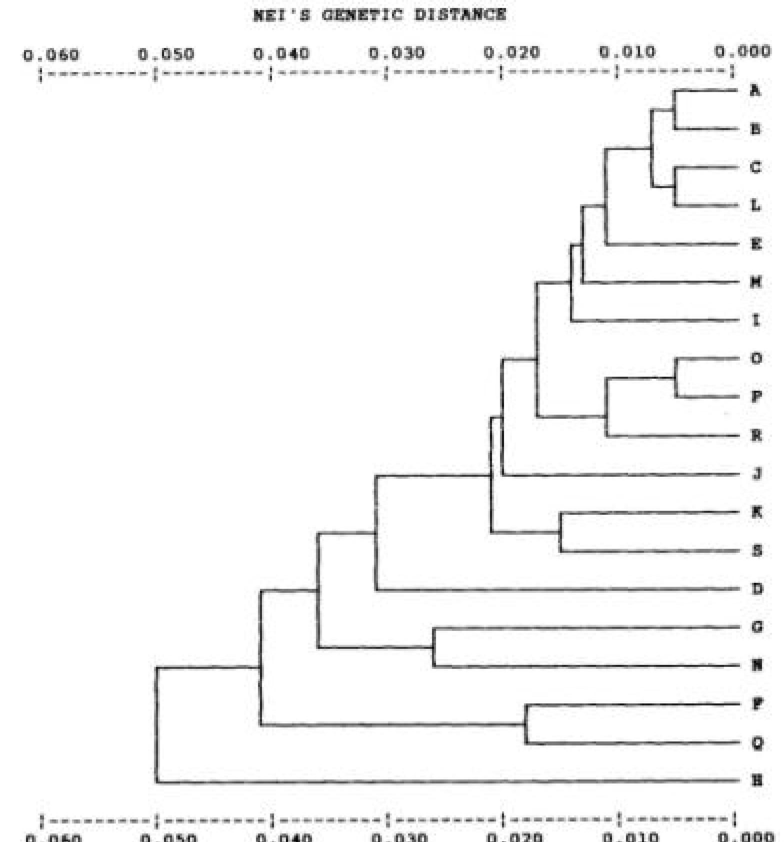


Fig. 2. Dendrogram from UPGMA cluster analysis based on Nei's (1972) genetic distance between the 19 populations of *Hemerocallis hakuunensis*.

The tree show demonstrates a significant intra species variation of the limited enzymes being assayed. It generally is consistent with what we have shown before. Thus one is led to assume that if one looked at the genetic variation across a large base of genes that the variation within species would be significant.

4.4.1 Morphological Classification

The classic approach to classification has been to use plant morphology. The use of such factors as those in the Table below has been done by many authors including those we have summarized in the introduction. Whether these are the best set are open to discussion.

Traits
No. of scapes (#)
Flower tube length (mm)
Petal length (mm)
Petal width (mm)
Stamen length (mm)
Pistil length (mm)
Plant (scape) height
Length of inflorescence minus flowers (cm)
Length of the lowest bracts (cm)
Number of flowers/scape (#)
Length of the perianth tube enclosing an ovary (cm)
Length of the outer perianth (cm)
Width of the outer perianth (cm)
Length of the inner perianth (cm)
Width of the inner perianth (cm)
Length of the widest leaf (cm)
Width of the widest leaf (cm)

We shall consider in detail the use of morphology in another paper. However it is worth considering what two people have done in the past one hundred years. In 1934 Stout published his book on daylilies. This was the first work and it was a work prepared by one skilled in the art. He was both a PhD in the field and he had even by then been active at the New York Botanical Garden, then and now a pre-eminent institution in the botanical area. He associated with Cronquist and others who had a great impact on the development of systematics. In his book he proposed a key to the species. It was a key, NOT a phylogeny. It was to be used to identify the species and NOT to specify any evolutionary or genetic relationship. He simply broke the species into two classes, those with branches and those without. Then he went down from there. Given what he had to work with, albeit extensive, he had not yet been able to identify all species and he did not have the advantage of thousands of others in the field.

In 1992 Erhardt in his book on Hemerocallis proposed a Classification, not a Key. The term classification carries a great deal more weight than a key. Keys help identify and classifications establish relationships. Erhardt states in his book:

"Stout's proposed division was not accepted and no one now supports it>"

Frankly that statement is a combination of arrogance and ignorance. By its face it uses the term division, not Key and not Classification. Division as a term of art has no standing. In addition if it was unused and in fact as implied by Erhardt was useless then why no one did from 1934 until Erhardt in 1992, sixty years, ever propose another, if we are to believe Erhardt. In fact there were dozens of others, all with slight nuances as new data was determined. Erhardt goes on:

"In my view there are five main groups of the day lily and the members in each group are either related or are perhaps varieties of one another."

Erhardt is a self declared "plantsman with wide ranging horticultural interests…" He clearly seems to lack the academic training given his self representation and one must ask what the basis for his selection was. There is no justification, just a statement of what he perceives as a fact. However, it is worth the exercise to examine his five morphological groups.

1. Fulva Group: Blooms are reddish, roots are bulb like, and this contains H fulva and H aurantiaca.

2. Citrina Group: Blooms mostly yellow, long perianth tubes and bloom opens in evening. The scapes are branched: H altissima, H citrina, H coreana, H lilioasphodelus (flava), H minor, H thunbergii.

3. Middendorffii Group: Blooms are orange, and they are sessile. Bracts are short and overlap: H dumortieri, H hakunensis, and H middendorfii.

4. Nana Group: Short scapes short perianth, not winter hardy: H forrestii, H nana.

5. Multiflora Group: Flowers on short stalks, branched, smaller flowers: H multiflora.

He then uses this classification to generate a Key. Thus he clearly knows or should know what the difference between a Classification and a Key is. Furthermore when he characterizes H hakunensis he says it is branched. Well it is or it is not. This is typical or Erhardt.

A better approach would be to look at the characteristics and see how they evolved. As we indicated earlier there is some evolutionary evidence to attest to the fact that citrina came from fulva. Then are all in the citrina group related, because of the night blooming. H minor is a very early bloomer, whereas citrina is later, typically four to six weeks, and coreana is even later. H altissima is very late and is a tall plant. They are all fragrant and one can hybridize between with some success.

In contrast H dumortieri and H middendorffii are both sessile, blooms at the same time, and seem to be cross sterile. Are these genetic variants of one another? On the other hand H hakunensis is a late bloomer and is not really sessile. Thus what was Erhardt's basis for making these aggregations?

We will examine this latter in the paper.

4.4.2 Genetic Classification

Recently, during the past twenty years, there has been a massive development in amount of tools available to both collect and analyze genetic data. Collection methods of proteins, enzymes, DNA, mRNA, cDNA, and variants of these have been developed. From the simple and now classic Southern blots to the use of million cell microarrays we now have a vast collection of raw genetic data potentially available.

The processing of this data for the genus Hemerocallis has just commenced. In addition to the collection techniques there are in many ways even more in terms of analytical tools. The tools range from the complex mechanism which align genes and search for genetic patterns, to those which take those patterns and use sophisticated statistical models to reverse engineer the evolutionary changes.

Techniques like the maximum likelihood technique have been used in communications for decoding signals which have been coded and sent across noisy and dispersive channels. The same processing used in this communications applications are now used in genetic analysis

4.5 Genetic Techniques

In this section we provide a summary of the key genetic techniques we need to understand in order to approach systematic from a genetic perspective.

4.5.1 Genes and Restriction Enzymes

This first section reviews several of the essential elements we need to take the next step and select genes and proteins. There are two elements we review; restriction enzymes and polymerase chain reactions.

4.5.1.1 Restriction Enzymes

One of the earliest discoveries in understanding DNA and genes was the recognition that certain enzymes, proteins, have the ability to cut DNA at certain well defined points in a consistent manner. These enzymes are called restriction enzymes and they allow one to select areas for cutting.

The following table is a list of some of the most important restriction enzymes. The Table lists the name of the enzyme, its source, namely what organism it has been obtained from, the target sequence it finds to cut at and the cut sequencing.

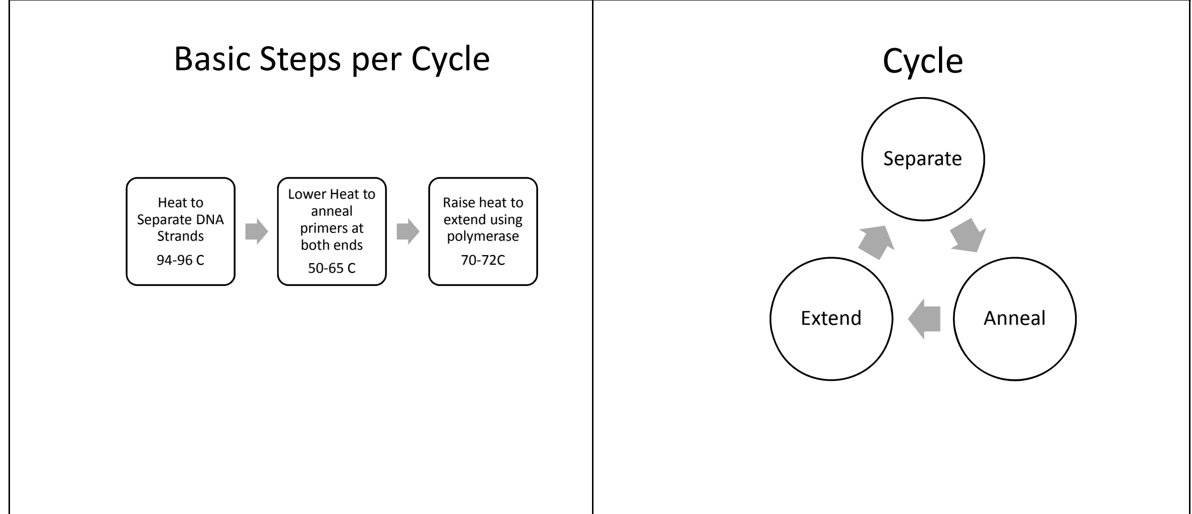
Restriction enzymes allow one to take long strands of DNA and to cut them in a predictable manner. Having these predictable cuts we can now add tags to the strands or do whatever else we seek to do.

Enzyme	Source	Recognition Sequence	Cut
EcoRI	Escherichia coli	5'GAATTC3'CTTAAG	5'---G AATTC---3'3'---CTTAA G---5'
EcoRII	Escherichia coli	5'CCWGG3'GGWCC	5'--- CCWGG---3'3'---GGWCC ---5'
BamHI	Bacillus amyloliquefaciens	5'GGATCC3'CCTAGG	5'---G GATCC---3'3'---CCTAG G---5'
HindIII	Haemophilus influenzae	5'AAGCTT3'TTCGAA	5'---A AGCTT---3'3'---TTCGA A---5'
TaqI	Thermus aquaticus	5'TCGA3'AGCT	5'---T CGA---3'3'---AGC T---5'
NotI	Nocardia otitidis	5'GCGGCCGC3'CGCCGGCG	5'---GC GGCCGC---3'3'---CGCCGG CG---5'
HinfI	Haemophilus influenzae	5'GANTC3'CTNAG	5'---G ANTC---3'3'---CTNA G---5'
Sau3A	Staphylococcus aureus	5'GATC3'CTAG	5'--- GATC---3'3'---CTAG ---3'
PovII*	Proteus vulgaris	5'CAGCTG3'GTCGAC	5'---CAG CTG---3'3'---GTC GAC---5'
SmaI*	Serratia marcescens	5'CCCGGG3'GGGCCC	5'---CCC GGG---3'3'---GGG CCC---5'
HaeIII*	Haemophilus aegyptius	5'GGCC3'CCGG	5'---GG CC---3'3'---CC GG---5'
AluI*	Arthrobacter luteus	5'AGCT3'TCGA	5'---AG CT---3'3'---TC GA---5'
EcoRV*	Escherichia coli	5'GATATC3'CTATAG	5'---GAT ATC---3'3'---CTA TAG---5'
KpnI[1]	Klebsiella pneumoniae	5'GGTACC3'CCATGG	5'---GGTAC C---3'3'---C CATGG---5'
PstI[1]	Providencia stuartii	5'CTGCAG3'GACGTC	5'---CTGCA G---3'3'---G ACGTC---5'
SacI[1]	Streptomyces achromogenes	5'GAGCTC3'CTCGAG	5'---GAGCT C---3'3'---C TCGAG---5'
SalI[1]	Streptomyces albus	5'GTCGAC3'CAGCTG	5'---G TCGAC---3'3'---CAGCT G---5'
ScaI[1]	Streptomyces caespitosus	5'AGTACT3'TCATGA	5'---AGT ACT---3'3'---TCA TGA---5'

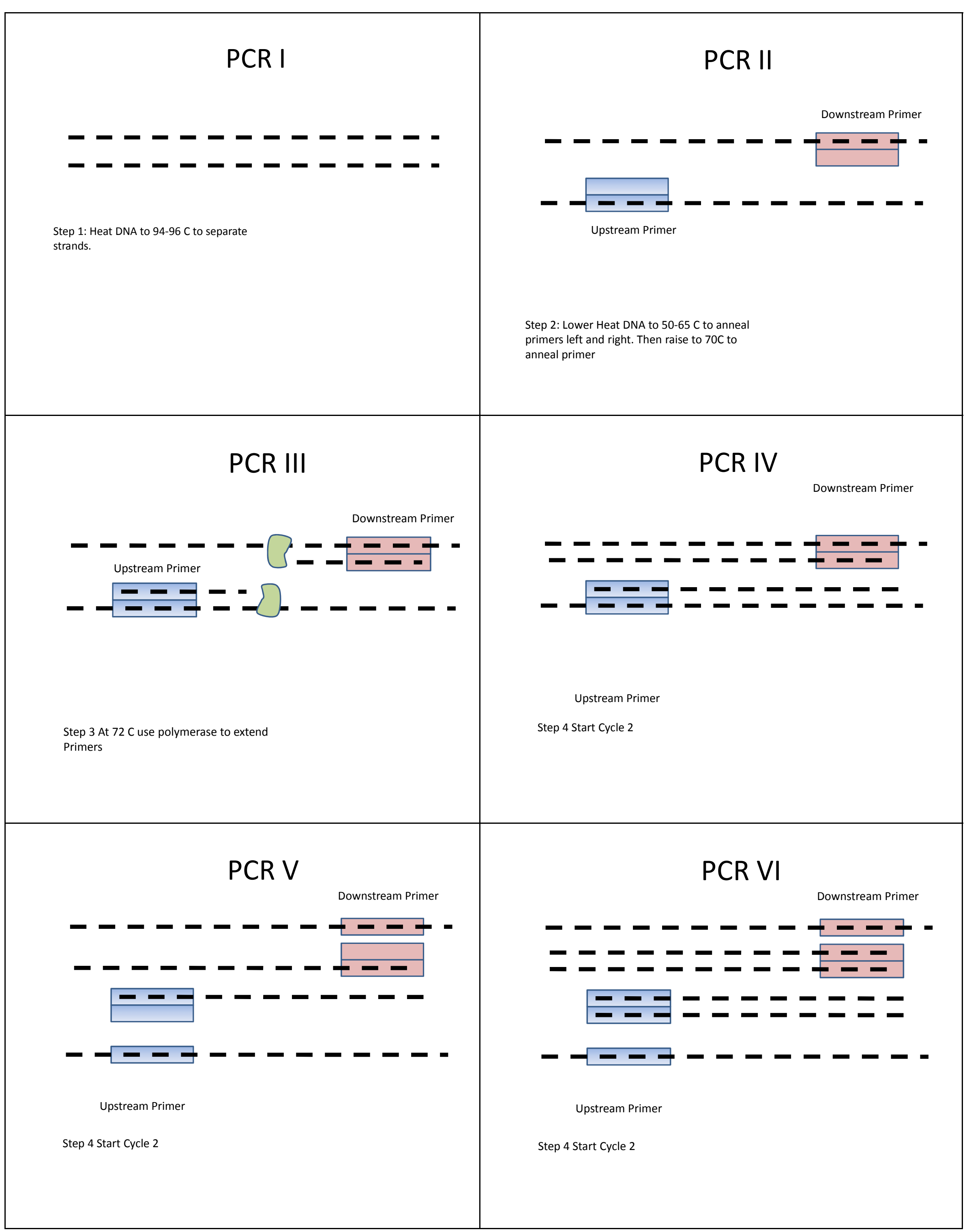
SphI[1]	Streptomyces phaeochromogenes	5'GCATGC3'CGTACG	5'---G CATGC---3'3'---CGTAC G---5'
XbaI[1]	Xanthomonas badrii	5'TCTAGA3'AGATCT	5'---T CTAGA---3'3'---AGATC T---5'

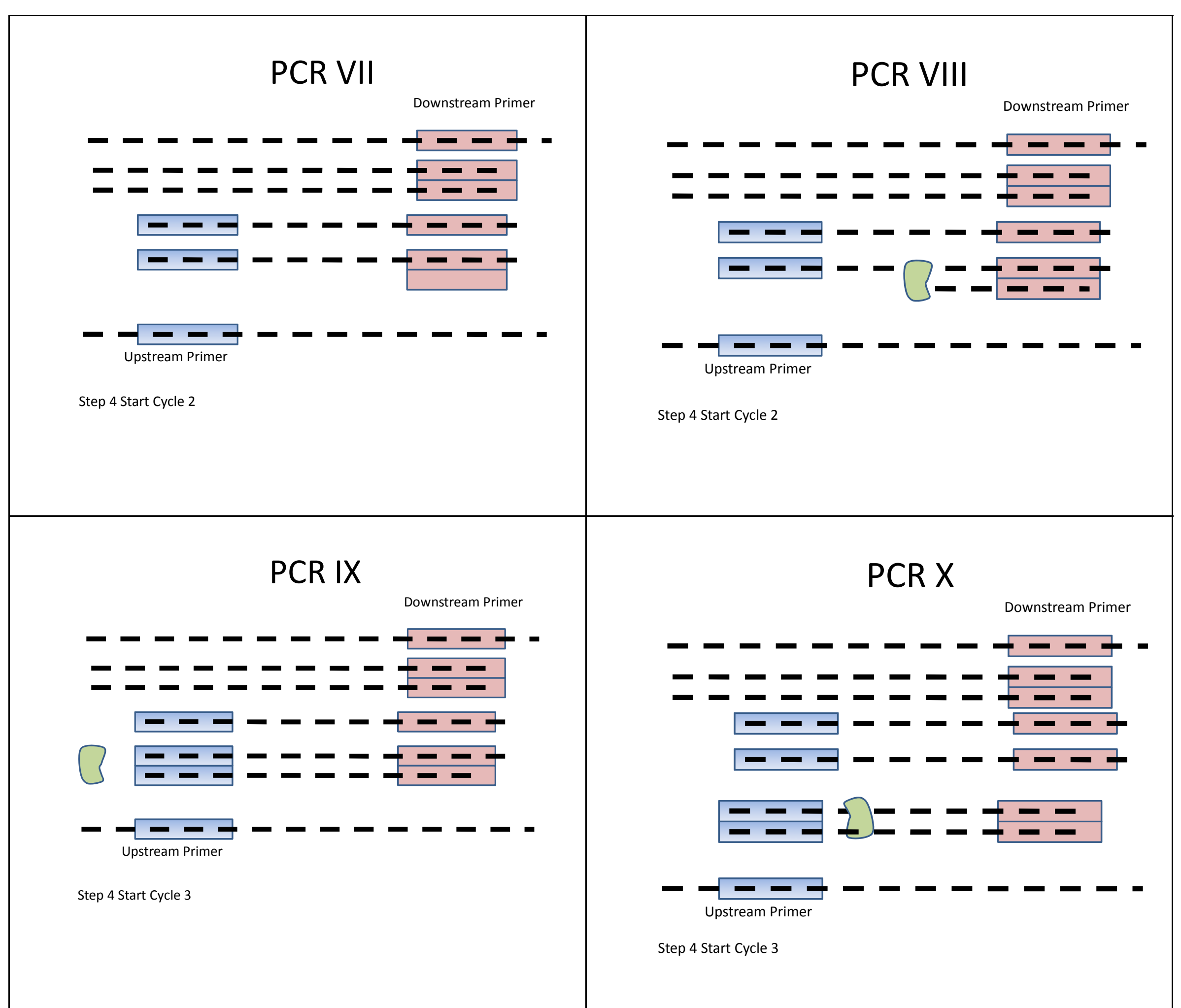
4.5.1.2 PCR

The polymerase chain reaction, PCR, was a brilliant step in the management of DNA. It allowed for the multiplication of small snippets of DNA into millions of copies of the small snippet. The process is shown at high level below. It goes through three heat stages, heat to break DNA apart, then cool to bond a marker, then heat again to get the enzymes to build out the DNA again along the new track created by the market. The separate, anneal and extend process is copied over and over.

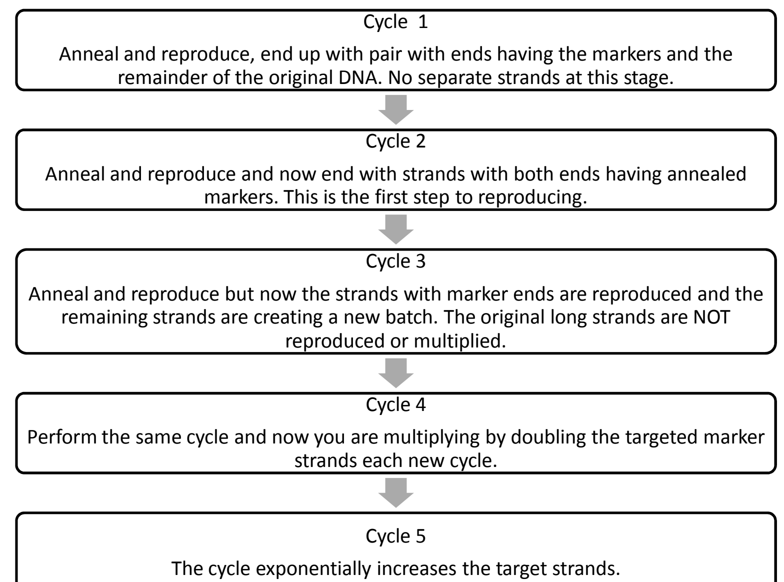


The specific details are shown in the following ten steps. Simply stated we separate, anneal a marker, re-grow the DNA now with the marker, repeat this with new markers, so that by the third step we have the segments with two end markers making themselves over and over, and they have exponential growth. Ten cycles, we get 2 to the 10th and this is a thousand fold multiplications, twenty cycles we have millions, all from a single strand!


PCR I
Step 1: Heat DNA to 94-96 C to separate strands.
PCR II
Downstream Primer
Upstream Primer
Step 2: Lower Heat DNA to 50-65 C to anneal primers left and right. Then raise to 70C to anneal primer
PCR III
Downstream Primer
Upstream Primer
Step 3 At 72 C use polymerase to extend Primers
PCR IV
Downstream Primer
Upstream Primer
Step 4 Start Cycle 2
PCR V
Downstream Primer
Upstream Primer
Step 4 Start Cycle 2
PCR VI
Downstream Primer
Upstream Primer
Step 4 Start Cycle 2



The following Table summarizes the cycles which we have shown above in extreme detail. The cycles require the first three to obtain a double ended segment and from that point on that specific segment is doubled at each part of the PCR cycle. There are also PCR systems which perform this cycling on a continuing basis.



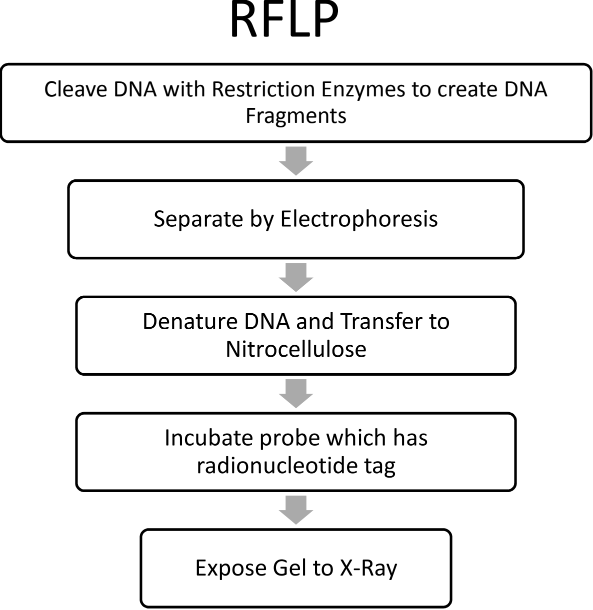
4.5.2 Procedures

We will now consider several procedures for the collection of genetically related data. Some use the basics of PCR and some do not. The AFLP approach which seems currently best for comparing species relies heavily upon PCR. The approaches we consider are as in the following Table.

Criterion	AFLP	RAPD	Microsatellite SSR	RFLP	Allozymes
Quantity of information	High	High	High	Low	Low
Replicability	High	Variable	High	High	High
Resolution of genetic differences	High	Moderate	High	High	Moderate
Ease of use and development	Moderate	Easy	Difficult	Difficult	Easy

4.5.2.1 RFLP

Restriction Fragment Length Polymorphisms or RFLP is one of the older mechanisms to obtain DNA fragments to analyze. The approach is detailed in the following Figure. Simply we use restriction enzymes then separate and use a probe to bind to the ends and then use an X ray which can detect the probe areas.



4.5.2.2 Microsatellite

Microsatellites are similar to the RFLP and instead of X-Rays we use fluorescent scans. The details of this approach are shown below. This is a small sequence approach of about six base pairs. Primers and PCR can be applied. The details are shown below.

Microsatellite

Isolate DNA

↓

Perform sequencing

↓

Design primers for regions flanking microsatellites

↓

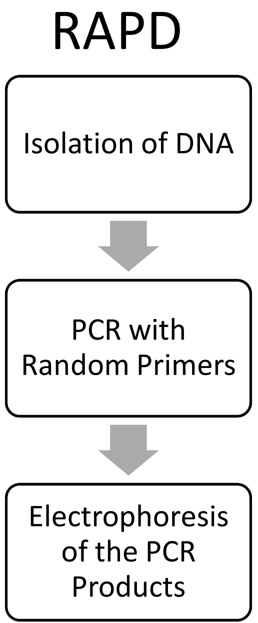
Electrophoresis separation of the amplicons

↓

Analyze by Fluorescent detector

4.5.2.3 RAPD

RAPD is Random Amplification of Polymorphic DNA. To some degree this is a "shot in the dark" approach. It generates many segments in a random fashion and then one can compare one species or plant to another. The mechanism is shown below in the Figure. This approach has not been used greatly in Hemerocallis analysis.

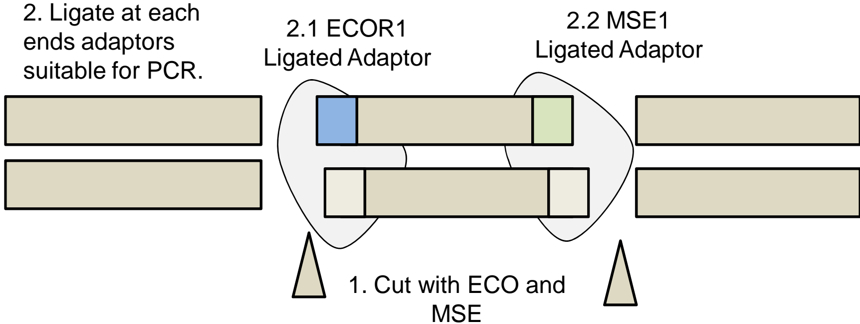


4.5.2.4 AFLP

AFLP or Amplified fragment length polymorphism is an intriguing approach which combines the best of all the other schemes. t gets long fragments, it has the ability to obtain quite a few and it has markers which give good results. It also uses PCR very effectively. The approach is shown in the following steps.

First we take DNA which we have extracted from the cell and then cut it with enzymes and after the cutting we ligate to the ends marker strips which we use to facilitate subsequent PCR.

AFLP I



Step 1: Take double stranded DNA and first, cut it with two enzymes, ECOR1 and MSE1. After having cleaved the DNA, then, second, ligate to the cuts the marker strips of DNA which can be used to facilitate PCRing the fragment.

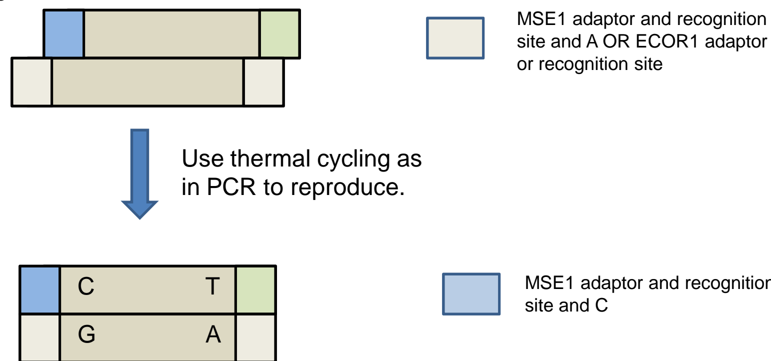
Reference: See *AFLP® Plant Mapping*, Applied Biosystems, 2007,

Second. we then use PCR to effect the growth of many small segments, typically many nucleotides long, and we can create a large amount of these segments. This is shown in the following Figure. This method is what is provided by Applied Biosystems.

AFLP II

Start with some pre-selection using the ligated primaries. Note that:

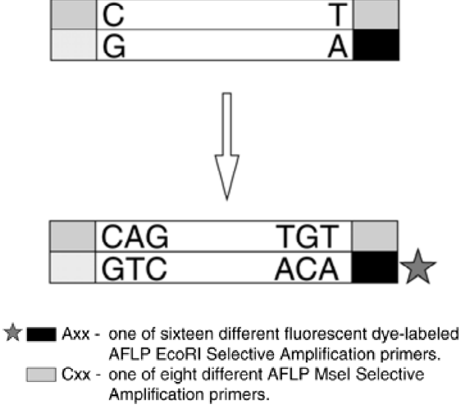
1. MseI the complementary primer has a 3´ C.
2. EcoRI the complementary primer has a 3´ A or no base addition.
3. PCR provided a preference in multiplication on the two end primed segments.
4. This process acts to purify the batch of segments to those with the two bases ligated.



Third, by using tagged primers on the ends of the fragments, we can use the 24 possible sets of primers obtain quite a large and diverse set of cuts. We do this both within a species and between a set of species.

AFLP III

This is a selective PCR process using tagged primers. The primer may be dye labelled and allows for selective processing. Additional PCR amplifications are run to further reduce the complexity of the mixture so that it can be resolved on a polyacrylamide gel. These amplifications use primers chosen from the 24 available AFLP Selective Primers (eight MseI and sixteen EcoRI primers). After PCR amplification with these primers, a portion of each sample is analyzed



Then we take the results and create electrophoresis results. This is shown below. In this Figure we show as rows bands of separate fragments which would result from performing an electrophoresis. Each column is a set of bands from a separate species. This simplified diagram shows how we can take many such fragments, from the possible 24 primers and the fact that each enzyme make cuts at different places, we get many possible fragments per plant. In the Figure the dark bands represent a fragment as it may possibly appear in an electrophoresis result. In addition the fragments are longer in the number of nucleotides so we can get a finer set of resolution than we could possibly attain in a single RFLP or RSS.

In addition, we can now use this data to create a set of relationships. Movement of bands means changes in genes, specifically nucleotides. Thus for small change we get a close match and for large changes we may get many splits. We then will use this data to create what we call a distance matrix which is a set of measure for showing how different species vary at the genetic level.

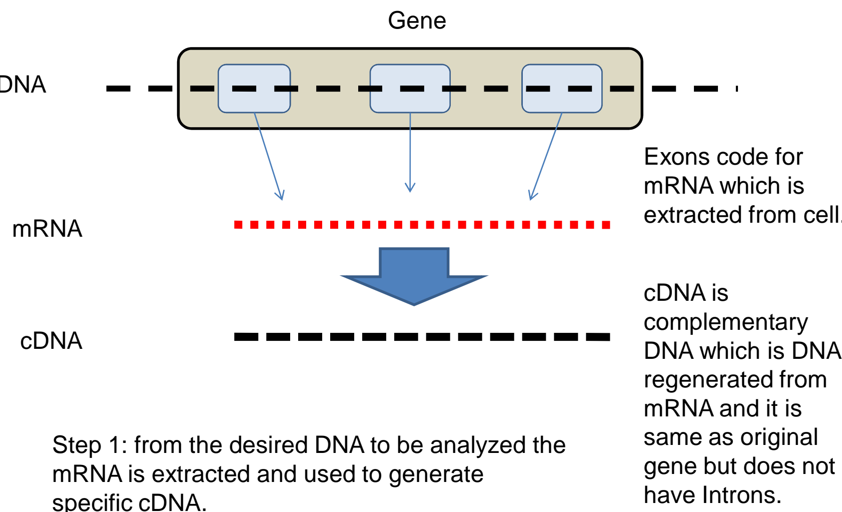
Band	Aurantiaca	Altissima	Dumortieri	Middendorfii	Fulva	Flava	Hakuuensis	Thunbergii	Minor
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									

4.5.2.5 Microarrays

Microarrays is a unique approach which allows for the analysis of millions of samples, it is a marriage of high tech solid state chip technology with DNA bonding. We describe it in the following four steps, each accompanied by a Figure.

Step 1: The first step in a micro array is the production of cDNA, or complementary DNA. cDNA is that set of nucleotides which account for the encoding of mRNA. It does not include the non-coding regions which are the introns.

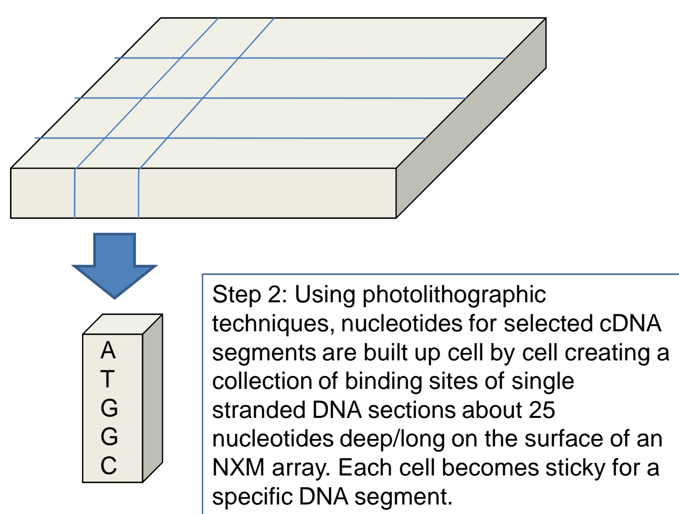
Microarray I



Step 2: In a separate environment we make the microcell. This is created in a manner identical to the making of integrated circuits which entails photo-masking techniques. Instead of silicon we used nucleotides. The array has millions of small holes in an array like manner. Each hole we fill with nucleotide, one nucleotide at a time.

For example, we can use the columns to drop DNA from each species sample and we then use each row with a set of 25 probe nucleotides to determine if that matching gene is present. The rows may be entered to match known genes, and using 25 sequential nucleotides we can fairly accurately get a gene. There are 4^{25} possible sequences and in Hemerocallis there are a few thousand genes, and we must know them otherwise we would be just "shooting in the dark". Microarrays do require knowledge of the CDNA library at least of key genes. We know, for example, from the work of Mol and Winkel Shirley the genes that control the secondary pathways for color. This we have discussed elsewhere.

Microarray II



Step 3: Now we take two DNA samples, one from what we call the Target, the plane we wish to categorize, and we use a reference plat as well, say H fulva. We then take the segments we collected in step one and tag then with green or red tags, green say for the Target and Red for the Reference.

Microarray III

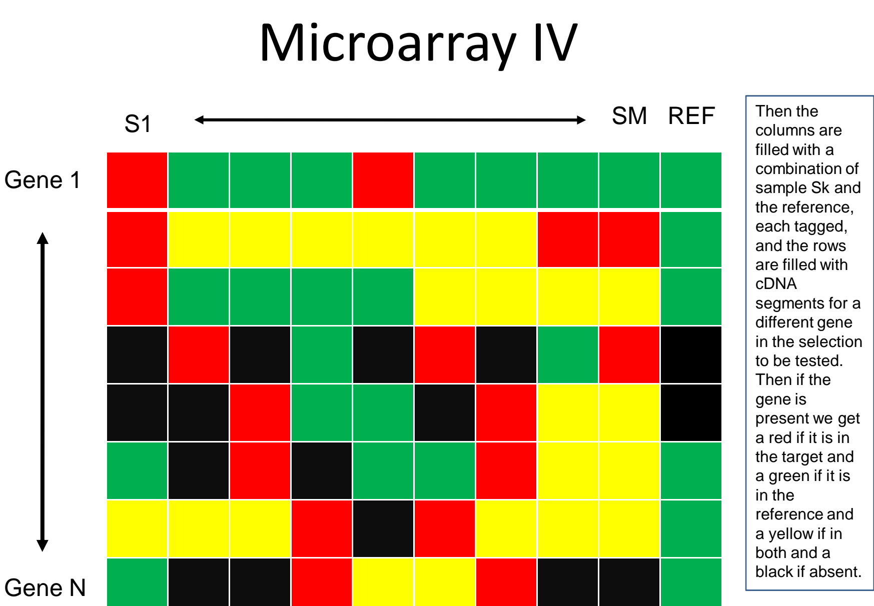
Target Cell

Reference Cell

cDNA

Step 3: For the DNA to be analyzed and a “Reference” target DNA, the mRNA is extracted from each and the cDNA is produced for every gene in the cells to be analyzed, and then it is tagged with a dye which is red for one and green for the other. Typically we tag the target red and Reference green.

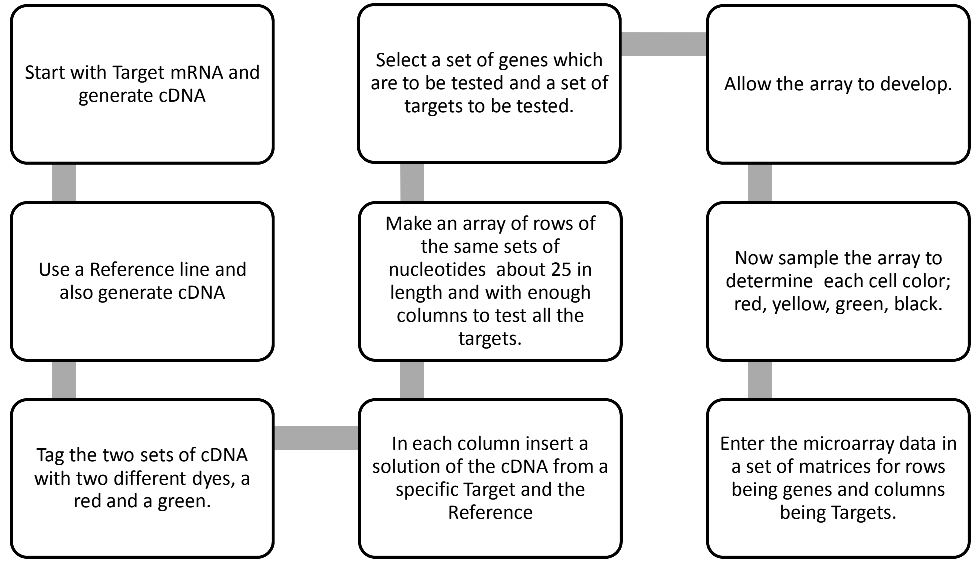
Step 4: We then take the samples from the differing plants, one in each column, and look at the array. If the microarray cell has the gene sequence we are seeking to march, and the Target has that sequence, it will bond and stick. If the Reference has it, it too will bond. If we just get the Target the cell will be green, if we just get the Reference the cell is red, if we get both the cell turns yellow, and if we have neither the cell is black. The result of a sample scan is shown below.



Now, we even get to try and look at the intensity of the red, green, or yellow. This we can try to see how much is expressed not just whether it is or is not. We will not discuss that here. In the above matrix we can see that many genes are expressed in one or both or none. If we have enough genes than we can argue we have the basis for an exceptionally good means to develop a classification.

In the following Figure we summarize the microarray process.

Microarray Summary



4.5.3 Comparisons

We can now compare the various methods we believe are effective. The three are the AFLP method, the second if the microarray and the third is total gene mapping. We defer the latter for the present.

	AFLP	Microarray	DNA
Advantage	Fast Can use many markers Can use NJ technique	Uses specific targeted genes Can provide for genetic variation with some time evolutionary analysis Can use NJ technique	Uses actual nucleotide sequences Can be used to determine time of evolution
Disadvantage	Limited number markers Does not reflect true genetic comparison Sequences are generally targets of opportunity	Requires known Genes	Requires large data sets Costly Analysis is complicated and should use ML techniques

4.6 CLASSIFICATION TECHNIQUES

We now discuss various classification techniques which use as input results from some form of DNA analysis such as the methods we have just discussed. Our goal is:

- Develop tools which can project relationships from data obtained using genetic material.
- Relate separate species to one another in a definable and metric based format.
- Look for consistency between gene based relationships and species based relationships.

4.6.1 PRINCIPLES

Trees are a graphical manner to represent relationships. The specific relationship we may wish to represent is one that reflects evolutionary relationships in time, namely which came first and which came after. In the development of trees using morphology we may look at sessile versus branched as a factor which may reflect temporal evolution. Namely in the monocots the sessile character may have some reason based upon paleobotany to have preceded the branching character. Thus, if we had such a basis or justification we would try to incorporate that factor.

The basic principles we try to use in developing trees are:

1. Parsimony: This is the Ockham's razor principle of using the simplest answer.

2. Bifurcation: New species come out one at a time because the enabling genetic change is one gene or one nucleotide at a time.

3. Time is reflected in a Distance Measure: There can always be create a distance measure between species based upon some set of characteristics. This measure may be morphological as having sessile versus branched, the length of a petal, the color of the flower, the width of a leaf. Or it may be genetic, the presence of an enzyme, the presence or absence of a specific sequence of nucleotides, or the number of restriction fragments across all chromosomes. The distance measure takes the difference and maps them to a number. The number then is related to them, namely how long does it take for a nucleotide to change.

4. Trees have inherent structures and the Classification meets the structure of the tree more than it may meet reality: We use the theory of trees to develop the dendrograms. This may limit what really happens.

5. Disbelief in "Black Swans": Black Swans are unexpected events. They may be a catastrophe, a crisis never expected, and some upheaval in nature. The models we generally develop assume the past acts continuously and that change is not the result of some dramatic change.

These principles make what we do in classification easier. We can model small and understandable change. We cannot model the unpredictable. Thus we should be aware that these models are the result of these assumptions and perhaps even more yet to be articulated. The classic systematics practitioner rationalizes what they do. One need look at the texts by Judd or those by Felsenstein, brilliant efforts and in many ways documents which reflect a reality. Yet they all lack the Black Swan which we know hides just in the shadows of all reality.

4.6.2 MEASUREMENTS AND METRICS

The first step is to define and develop metrics, distance reflective of change, measures that map measurements of reality into manipulateable numbers. In our analysis we will focus on data measurements resulting from the techniques we detailed in the previous section. As such we can look at three general areas:

1. Gene Dynamics (Nucleotide Changes of ATGC): This may use the classic Jukes Cantor measure of change of nucleotides which assumes equal probability of nucleotide change per unit time. We may measure the nucleotide strings, nucleotide by

nucleotide and from this try to see how they may be best arranged so that we may characterize evolution consistent with a model of change based upon some reality.

2. Inferred Genetic Distances: This approach uses data such as AFLP data and the like and then defines a distance between them in some manner which reflects gene change. In our case we use 1, 0 as binary. Could also use measure of number of nucleotide changes if that could be determined. Microarray data could be used here as well.

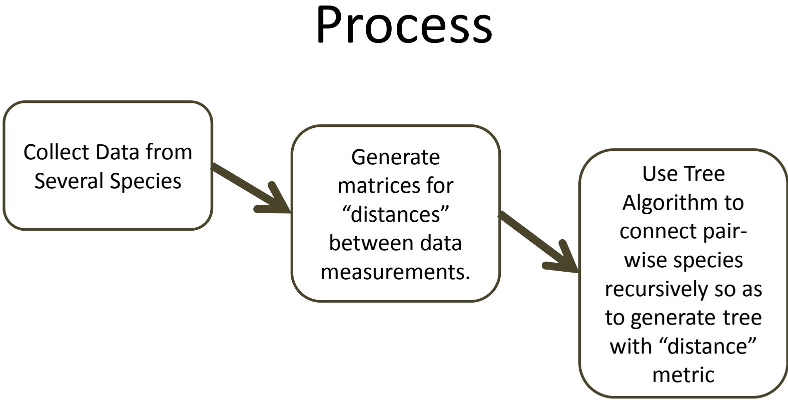
3. Non Genetic based upon ODU. These may also be clustering techniques and it does not utilize measurements of the type we look at here.

The overall process which we are to follow is depicted in the following figure. It is a simple three step process:

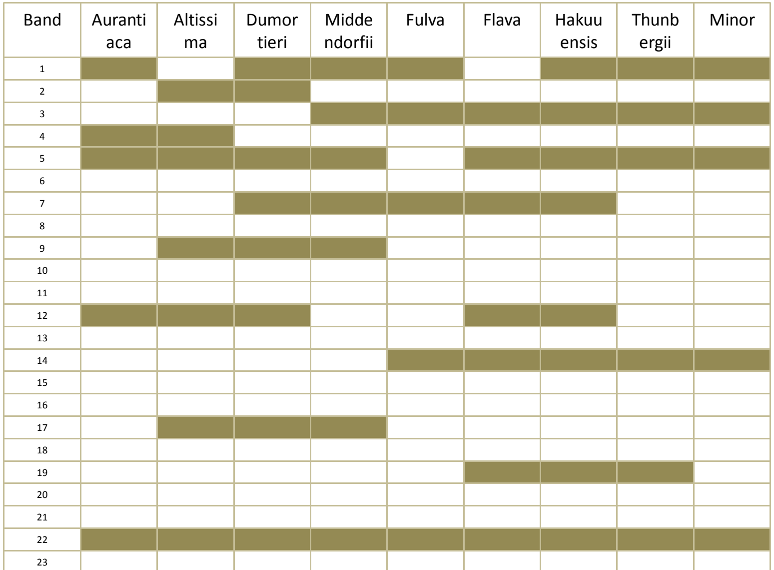
1. Obtain the raw genetic data.

2. Create distance measure based upon the data.

3. Employ the distance measure to create trees.



The following is a repeat of the AFLP data that we may collect in an experiment. It is a rendition of an electrophoresis plot of the AFLP marker sequences. We take this chart can convert it to a distance matrix.



AFLP Data

Band	Aurantiaca	Altissima	Dumortieri	Middendorfii	Fulva	Flava	Hakuuensis	Thunbergii	Minor
1	■		■	■	■		■	■	■
2		■	■						
3				■	■	■	■	■	■
4	■	■							
5	■	■	■	■		■	■	■	■
6									
7			■	■	■	■	■		
8									
9		■	■	■					
10									
11									
12	■	■	■			■	■		
13									
14					■	■	■	■	■
15									
16									
17		■	■	■					
18									
19						■	■	■	
20									
21									
22	■	■	■	■	■	■	■	■	■
23									

The following depicts a distance matrix for this AFLP data. Let us assume we have M species and N measurements for each species in the electrophoresis chart. Thus the above has M columns and N rows. We then create a matrix which measures the distance between two species. It is as below. We look at the two columns and we generate a distance as the number of electrophoresis bands where they differ.

We define the distance measure as follows:

$$D_{i,j} = \frac{1}{N}\sum_{k=1}^{N} d_{i,j}$$

$$d_{i,j} = \begin{cases} 1 \text{ } if \text{ } there \text{ } is \text{ } a \text{ } band \text{ } in \text{ } one \text{ } and \text{ } not \text{ } the \text{ } other \\ 0 \text{ } if \text{ } there \text{ } is \text{ } a \text{ } band \text{ } in \text{ } both \text{ } or \text{ } no \text{ } band \text{ } in \text{ } both \end{cases}$$

The following shows the distance matrix between the species. Again this is a repeat of details we presented earlier. Note that the matrix is symmetric.

Distance Matrix

	Aurantiaca	Altissima	Dumortieri	Middendorfii	Fulva	Flava	Hakuuensis	Thunbergii	Minor
Aurantiaca	0	1	3	4	7	2	9	3	5
Altissima	1	0	4	7	9	2	3	5	6
Dumortieri	3		0	3	6	7	9	3	2
Middendorfii	4			0	2	6	9	3	5
Fulva	7				0	5	2	3	9
Flava	2					0	8	2	5
Hakuunensis	9						0	2	9
Thunbergii	3							0	4
Minor	5								0

There are a few issues we must be concerned with. First are the options of a measure. For example: (i) develop binary measures such as {0,1} values based expression or non-expression of gene and (ii) create artifact distances as a measure expression by measuring the density of the color; thus a variable on the interval [-1,1]. Second there are many issues that need to be focused on such as: (i) Sensitivity of the measurements, (ii) Use of a reference mix and (iii) All issues related to errors in microarrays and their measurements.

4.6.3 Techniques for Trees

The following are the principle techniques found in the development of Trees:

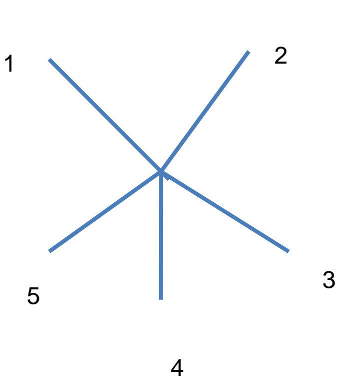
- Neighbor Joining: Tries to get a tree with the best possible fit of an additive rooted binary tree.
- Maximum Likelihood: Assumes an underlying transition process and then attempts to create a tree based upon a best fit to that process.
- Maximum Parsimony
- Generalized Neighbor Joining
- Weighted Neighbor Joining
- Un-weighted Pair Group with Arithmetic Mean (UPGMA)
- Minimum Evolution
- Fitch-Margoliash-Least Squares Fit

We will look at two of these; Neighbor Joining and Maximum Likelihood.

4.6.4 Neighbor Joining

The Neighbor Joining scheme is used frequently. It was developed in the mid 80s and was modified to correct initial errors in the analysis and also to improve the running time of the algorithm.

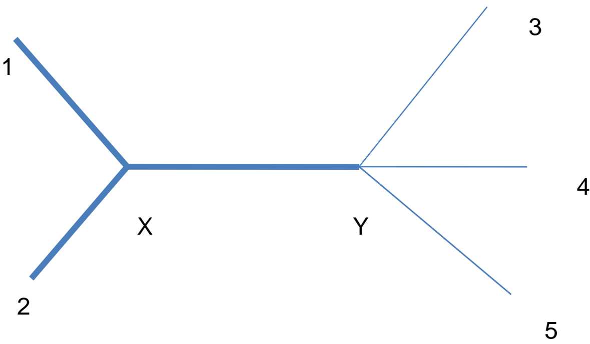
Tree Generation I



Start with say 5 of the possible species. They are all connected and we know we have a distance matrix which give a "distance" between all pairs of this collection. We now want to create a "tree" on a pair-wise basis so that there is some sound relationship between the end points, namely the species.

Then we begin the development of a tree. This we depict below. Inherent in this process is the assumption that species split from a common ancestor in pairs. Namely we have a binary set of nodes; we never get three species at a split, only two.

Tree Generation II



We start the tree process by selecting in some manner pairs of “closest” end points and then building this out.

We focus on Trees which are additive Trees. A tree is a connected graph which has no cycles. In a tree there is a unique path between every pair of vertices. An Additive Tree is a tree which has certain properties. Namely in an additive tree we have:

$$L_{ij} = \text{length of any path in the tree between any two points.}$$
$$L_{ij} = L_{in} + L_{n,n+1} + \ldots + L_{N,j};\ \text{where each are lengths of contiguous segments}$$

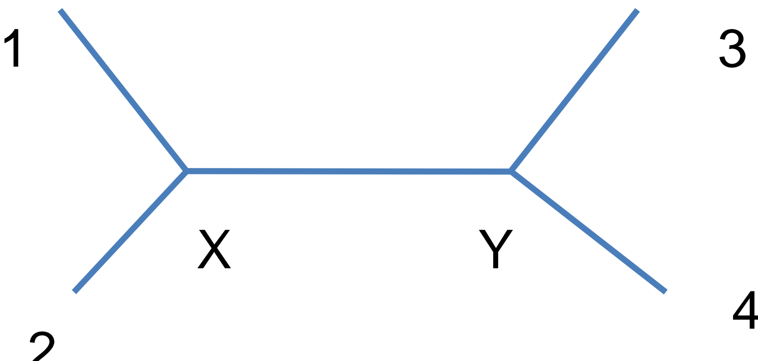
Thus in our example we have:

$$D_{1,2} = L_{1,X} + L_{X,2}$$

A binary tree is a tree with bifurcated ends, namely only two vertices at each branch in a tree. Finally a rooted tree is a tree is a binary tree with a single starting point reflecting evolutionary trends.

Consider the following simple Tree.

Additive Tree



We can show that the entries in the distance matrix and the path lengths can be calculated.

$$D_{12} = L_{1X} + L_{X2}$$
$$D_{34} = L_{3Y} + L_{Y4}$$
$$D_{13} = L_{1X} + L_{XY} + L_{Y3}$$
$$D_{23} = L_{2X} + L_{XY} + L_{Y3}$$
$$D_{24} = L_{2X} + L_{XY} + L_{Y4}$$

And we can write this as:

$$\begin{bmatrix} D_{12} \\ D_{13} \\ D_{14} \\ D_{23} \\ D_{24} \\ D_{34} \end{bmatrix} = \begin{bmatrix} 11000 \\ 10101 \\ 10011 \\ 01101 \\ 00110 \end{bmatrix} \begin{bmatrix} L_{1X} \\ L_{2X} \\ L_{3Y} \\ L_{4Y} \\ L_{XY} \end{bmatrix}$$

We can now state the Neighbor Joining Algorithm:

1. We begin with all of the vertices in a star formation and then we compute for each pair of vertices the factor S_{ij}. We select the pair with the least value. Recall:

$$S_{12} = L_{XY} + (L_{1X} + L_{2X}) + \sum_{i=3}^{N} L_{iY}$$

2. Using the relationships between the L and D elements we can write this as:

$$S_{12} = \frac{1}{2(N-2)} \sum_{i=3}^{N} (D_{1k} + D_{2k}) + \frac{1}{2} D_{12} + \frac{1}{N-2} \sum_{j=3}^{N} \sum_{3 \le i}^{j-1} D_{ij}$$

3. Calculate this for all pairs and select the pair to join which has the smallest S value.

4. Now we have a tree with N-1 vertices But the new vertex is the combination of 1 and 2, we call it X. We now need to obtain the new D values that relate to this new vertex. We define that as:

$$D_{xj} = \frac{\left(D_{1j} + D_{2j}\right)}{2}; (3 \le j \le N)$$

5. We then go back to step 2 and use these new values and select again the new pair that gives the smallest S value. We repeat this process until we have all pairs.

6. The dendrogram is the result using NJ and AFLP data for the various Hemerocallis species.

4.6.5 Maximum Likelihood

Maximum Likelihood is an approach to classification using genetic data, genes specifically, and it incorporates details about the changes in the genes over time. The maximum likelihood approach assumes that we have obtained a mapping of the gene or some gene segment down to the nucleotide. Then it assumes we have the same segments for the other species we wish to compare. Let us assume we have twelve species and we have the following twelve 25 nucleotide long segments. We can assume that they come from a cDNA, recalling that cDNA is made from mRNA using a reverse transcriptase. Thus we have the following as in the Table:

Species	*cDNA Segment*
altissima	**AATTCTACTTACTTACTGGACCAGT**
aurantiaca	**AATTCGGCTTGCGTACTGGACCAGT**
citrina	**AATTCCCCTTACTTACTGGACCAGT**
coreana	**AATTCGGCTTACGCGCTGGACCAGT**
dumortierii	**AATTCGGCTTACTTACTGGACCTAA**
flava	**AACGCGGCTTACTTACTGGACCAGT**
fulva	**AATTCGGCTTTAATACTGGACCAGT**
hakunensis	**AATTCGGCGGACTTACTGGACCAGT**
middendorfii	**AATTCGGCTTACTTACTCCACCAGT**
minor	**AATTCGGCAAACTTACTGGACCAGT**
multiflora	**CCTTCGGCTTACTTACTGGACCAGT**
thunbergii	**AATTCGGCTTACGGACTGGACCAGT**

We may hypothesize that the original sequence is **AATTCGGCTTACTTACTGGACCAGT.** If we did then we have noted the changes in the sequences by the red nucleotides in each of them. We then pose the following problem:

1. Assume we have 12 nucleotide sequences from a segment of cDNA we know to be a useful segment in determining a plant characteristic, such as color.

2. For each of the segments, that is, for each species, define a vector of dimension 25X1 as z(n) for each of the 12 species.

3. Assume the following:

   a) There existed a common ancestor for all of these species.

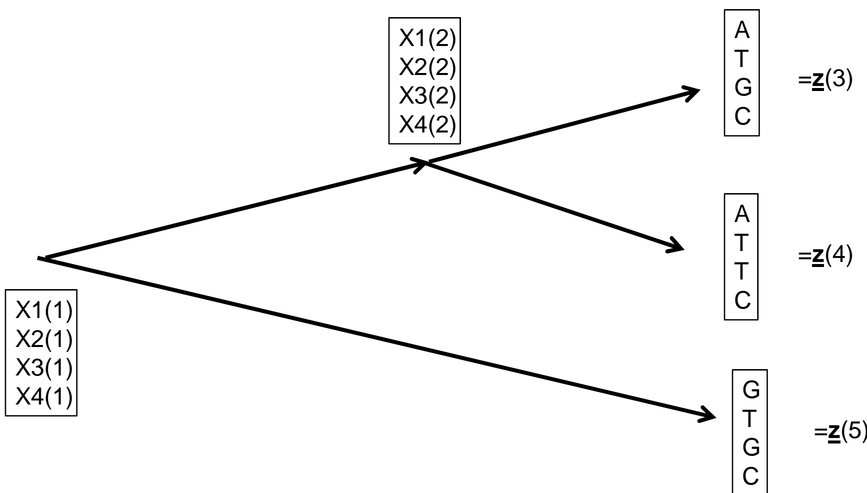
   b) Evolution occurs at one nucleotide change at a time and is binary. Namely we do not get multiple nucleotide changes and we do not get binary change happening simultaneously.

   c) Assume that we can ascribe a probability to a single nucleotide change, and we may or may not know the value and the value may or may not remain constant over the time horizon.

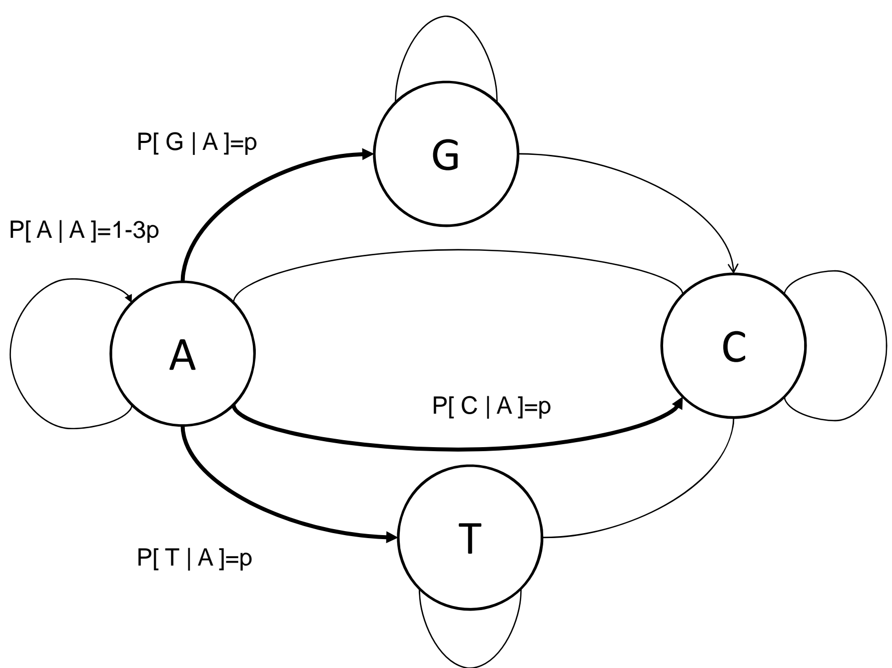
Thus with these assumptions there will exist a single rooted tree for this set of species.

4. The changes that occur do so independently. That is we have a Markov process.

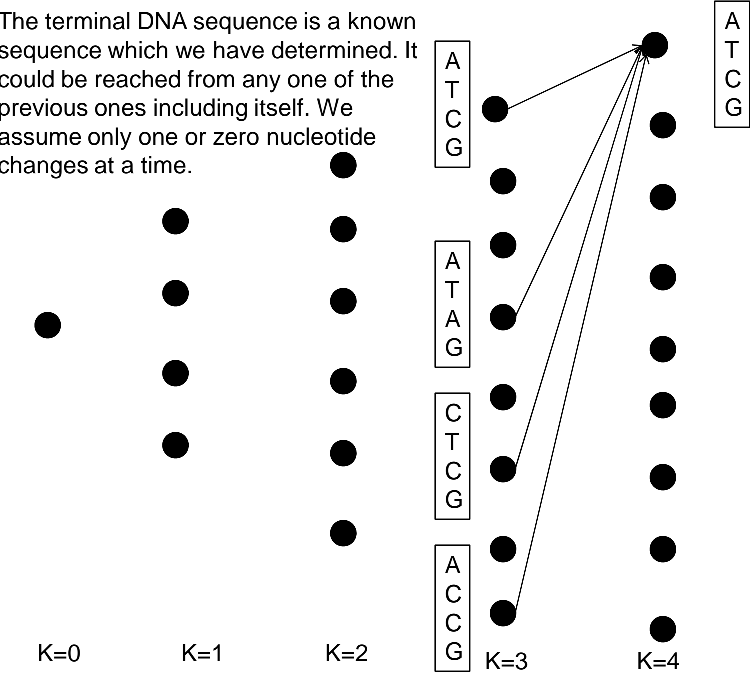
The following Figure depicts what we are posing. The internal nodes, assumed to be 25 nucleotide sequences also are labelled as x(n). They are 25X1 vectors as well



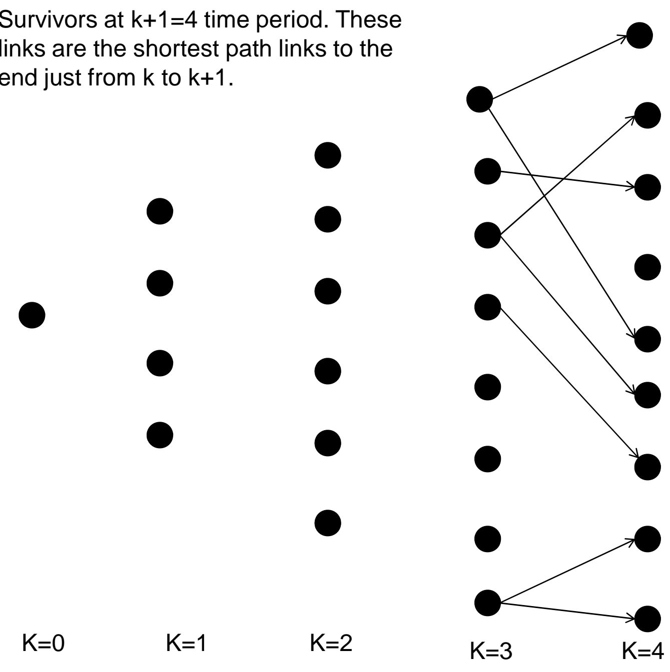
We can model the nucleotide changes as a Markov process and we can use a finite state machine to do so. This we show below. The probability of changing a nucleotide is p and is the same for all changes. This is the simplest model possible. One may look at various other and more complex models.



The problem can now be further posed as shown in the Figure below. We have the end points on a sequence of changes. We know the limits we have placed on the changes and we now want to find a process which will give us the "best" set of past changes so that we get what we observe in the 12 different 25 nucleotide sequences. In addition we want to get a single end point tree which is generated by single bifurcations. The following Figure looks at the end point.



if we just look at the last steps, we know that if we had some algorithm which gave us the best path then there would be some best path to every one of the know end elements. Then we would ask how we got to them. This end element best path is shown below.



Now the principles of the maximum likelihood approach are:

- Deals with DNA Sequences
- Known rates of nucleotide change per unit time
- Changes result in two new paths and no more at any one time
- Changes always go “forward”, no crossing or reverses

Further

- Assume we have ACTG type nucleotides
- Assume that there is a rate of change of α per unit time. Thus over T units of time the probability of a single nucleotide change is p which is αT
- Assume all are equally the same
- Then we have a finite state machine model for the change

The problem is then to find the sequence of change states which lead to the known final states so that the sequence maximizes the a posteriori probability or as in the following:

$$\max\ p(x(1)...x(n)|z(1)...z(m)) = \max \frac{p(z|x)p(x)}{p(z)}$$

This is the maximum likelihood approach. Let us explain it a bit.

1. The probability density, p(x|z) is the a posteriori probability of some or all of the internal nodes, we call them x, give the observed end nodes, and we call them z.

2. We want to find the set of all possible internal nodes, the set of all possible xs, that can yield the observed z, and we want that specific set of x which maximizes the a posteriori probability. Well one may ask why that is a good thing to do. There have been many analyses of this problem but the best approach is looking at detection of targets in radar, where this was most effectively used. Selecting this point maximizes the target hit probability and minimizes the false alarm rates.

But we can also write the above in terms of the p(z|x) and then the p(x). We can reject the p(z) since it has no impact on choosing the x.

Now since we have structured this with Markov processes, and since this means that changes depend only on their immediate past we can write:

$$p(z \mid x) = p(z(1) \mid x(k))\, p(z(2) \mid x(j)) ... p(z(m) \mid x(r))$$

That is we can write the $p(z \mid x) = p(z(1) \mid x(k))\, p(z(2) \mid x(j)) ... p(z(m) \mid x(r))$

And recall that we have:

$$p(x(j) \mid x(k)) = \begin{cases} 1-3p \\ p \end{cases}$$

Thus in our earlier initial map with end nodes we can write for each the following:

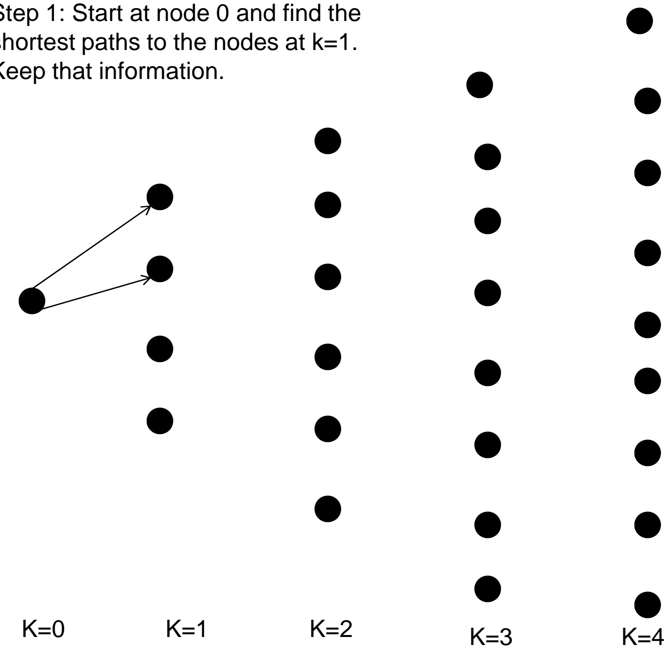
$$\begin{aligned} p(z(3) \mid x(2), x(1)) &= p(z(3) \mid x(2))\, p(x(2) \mid x(1)) \\ p(z(4) \mid x(2), x(1)) &= p(z(4) \mid x(2))\, p(x(2) \mid x(1)) \\ p(z(5) \mid x(1)) &= p(z(5) \mid x(1)) \end{aligned}$$

We find it more convenient to define a distance value defined as:

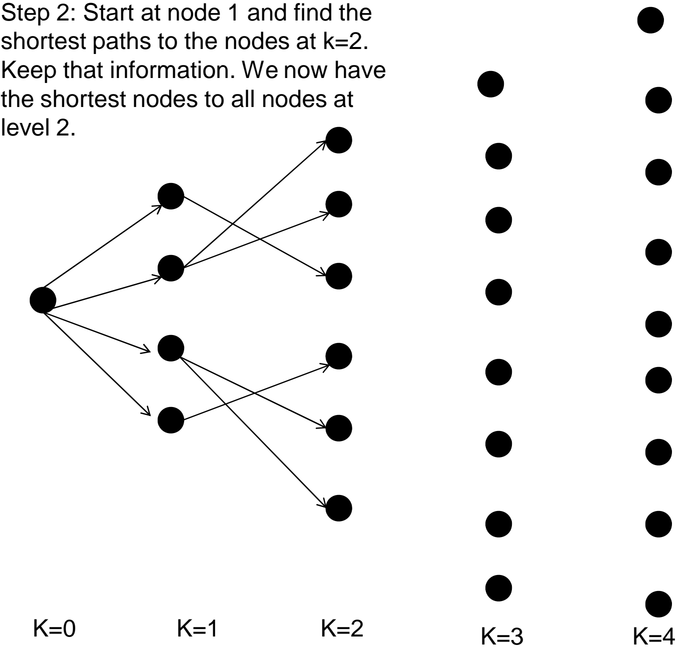
$$\lambda(\zeta(k)) = -\ln p(x(k+1) \mid x(k)) - \ln p(z(k) \mid x(k))$$

Thus instead of maximizing the probability we minimize the distance as defined above. In addition we can perform computations better this way. Thus we are seeking the minimum length path through the network where we define length as above. This is called the Viterbi algorithm and was used first in decoding convolutional codes.

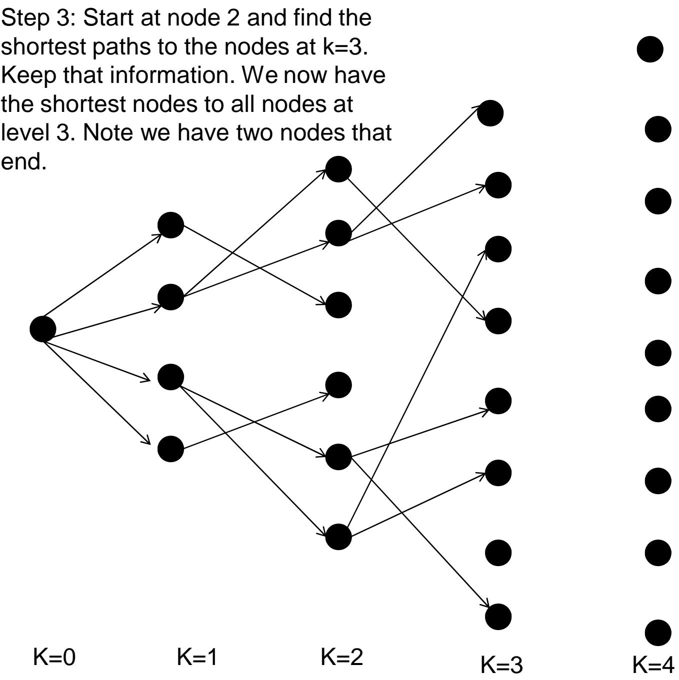
The algorithm is developed graphically as follows:



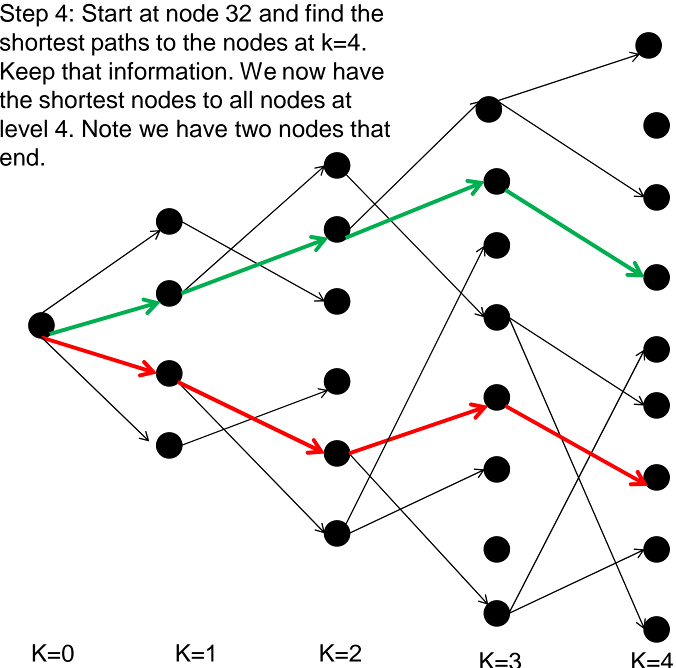
Thus step 1 starts at the beginning. Here we assume the beginning for some node. We can and will do it for all possible nodes. Recall that for a 25 nucleotide sequence we have 4^{25} nodes. Then we go to step 2 as below.



In the above step we start culling out nodes. We keep only at k-2 those paths of least length where we define length as above. We continue as below:



Finally we get to the ends, and each path to each end node is the least distance path.



Note several things about this algorithm.

1. All paths are minimum length from the selected initial point. If we change the initial point we get a whole new set of paths. To determine the best initial point we do this for all possible initial points and choose the one with the least sum of the lengths. This is computationally intense.

2. The red and green paths above show details of two specific paths. Note in both there is no branching at step k=3. The other end points branch at k=3. This we have a binary tree perforce of the assumptions and not a result of anything we see in the data. The assumptions are control the end result often more than the data so beware assumptions.

3. The resulting tree becomes evident. One need just follow the path.

Many authors have tried to explain this approach to no avail. I have seen such works as that of Durbin et al which make it totally incomprehensible! One should beware those who have notation which is incomprehensible.

4.7 Application to Hemerocallis

The first extensive efforts at taxonomy within the daylily were attempted by A.B. Stout (1934), in which two major classifications were proposed: those having branched scapes (Euhemera) and those without branched scapes (Dihemera). Stout's classification,

however, is now not generally well accepted. A more recent classification of daylily species into five major groups is presented by Erhardt (1992), and generally supported by the AFLP data in the present study. Erhardt's classification of the five groups comprises (1) *fulva, (2) citrina, (3) middendorffii, (4) nana, and* (5) *multiflora.*

Utilizing neighbor-joining analysis, the six *H. fulvas* were distinctly separated from the other species. Clustering within the *fulvas* also supported some fine-scale taxonomic classifications. For example, the distinction described by Erhardt between the two fulva double-flowered genotypes 'Kwanso' and 'Flore Pleno' is reflected in the molecular data. Within the *middendorffii* group, *H. dumortierri*, *Hemerocallis middendorffii* and *Hemerocallis hakunensis* all grouped together as proposed by Erhardt.

However, the distinction between the *citrina* group and the *middendorffii* group was not well defined and contained some overlap. *H. citrina* and *Hemerocallis minor* were grouped together as proposed by Erhardt, but were also grouped with members of the *middendorffii* group. Erhardt had proposed a close relationship between two other members of the citrina group, *Hemerocallis lilioasphodelus* and *Hemerocallis thunbergii*, which was well supported by our data, but they did not closely group with the other *citrina* members. In fact, our data suggest that the *middendorffii* group and the *citrina* group should be merged into one large taxonomic group.

The only major anomalies among the species analysis were supposed clonal variants of *H. citrina* (var. Vespertina) and *Hemerocallis dumortierii* (var. Sieboldii). While both did cluster within the *middendorffii-citrina* group, they did not closely group with their respective parental clones from which they were supposedly derived. Traditionally, there have been a number of variants of *H. dumortierrii* in commerce.

Thus, the variety Sieboldii may or may not include the traditional species *H. dumortierrii* as a direct ancestor even though there are phenotypic similarities. *H. citrina* is self-incompatible and thus any variant arising from it would have to be obtained from an outcross. Hence, these genotypes may either have arisen via cross-pollination or may represent distinctly different genotypes. The following Table recounts Tompkins et al AFLP data.

Tomkins and his team performed analyses on the dozens species and hybrids, a massive number but readily doable with AFLP. The following Table depicts the targeted species and the year they were identified.

Genotype	*Year*
H. citrina	1897
H. citrina var. Vespertina	1941
H. dumortierii	1830
H. dumortierii var. Sieboldii	Unknown
H. fulva Europa	1762
H. fulva Flore Pleno	1860
H. fulva var Kwanso	1860
H. fulva var Maculata	1895
H. fulva var Rosea	1924
H. fulva var Sempervirens	1966
H. hakunensis	1943
H. lilioasphodelus	1576
H. middendorffii	1860
H. minor	1748
H. thunbergii	1873

The AFLP primers used on these species and the hybrids are shown in the following Table.

Primer combination	*Total number of bands*	*Polymorhic bands*	*Polymorhism (%)*	*Scored bands*
E-AAG/M-CAA	126	93	74	None
E-AAG/M-CAC	135	109	81	None
E-ACC/M-CAA	130	109	84	None
E-ACC/M-CAC	103	84	82	61
E-ACC/M-CAG	87	66	76	36
E-ACT/M-CAT	136	108	79	None
E-ACT/M-CTT	107	84	78	55
E-ACT/M-CTA	82	63	77	None
	Total 906	Total 716	Mean 79	Total 152

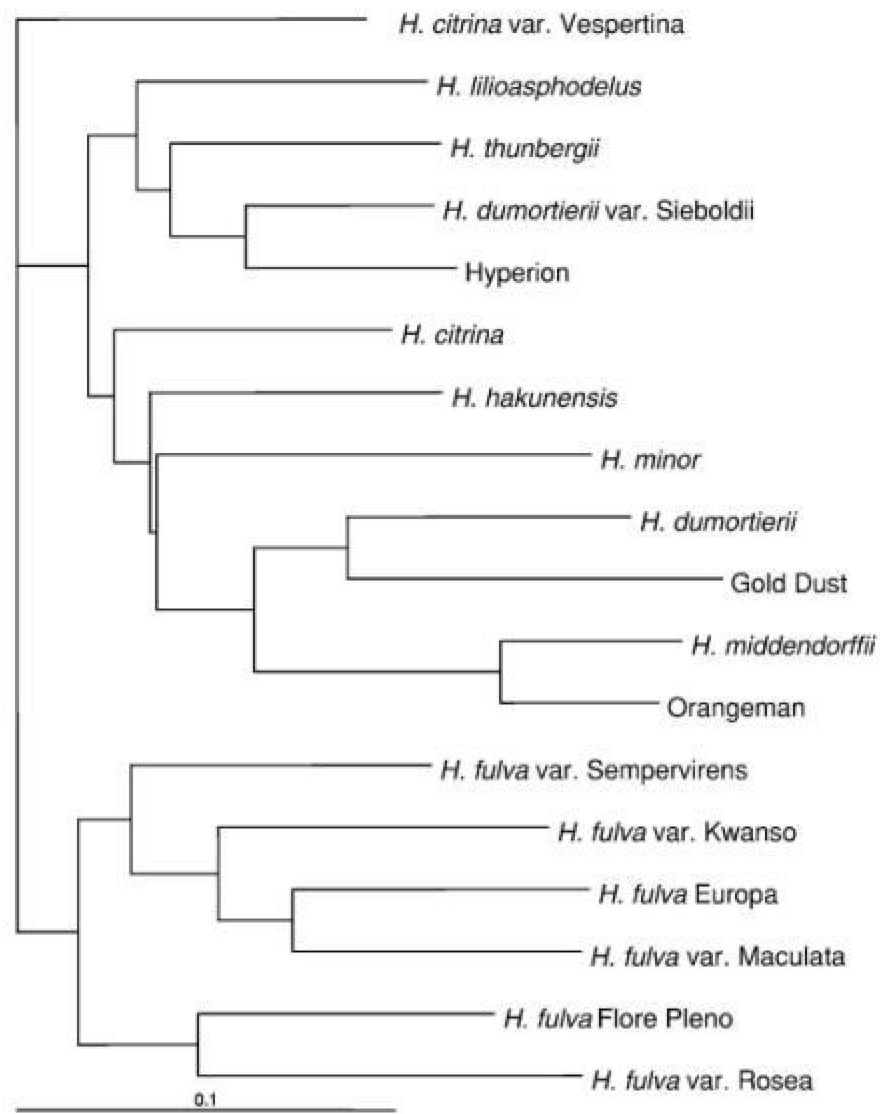
Using this data and employing the NJ technique, Tomkins et al have obtained the following dendrogram. It is effectively a classification using the AFLP markers. The following observations can be made:

1. The species identified as H citrina Vespertina is ranked as the initial root species.

2. H fulva (Europa) the triploid variety is shown to be an offshoot from the split with H fulva kwanso.

3. H citrina as species is an offshoot of Vespertina and is aligned with minor, hakunensis and dumortieri. This seems a bit strange.

4. The H fulvas are all grouped together.



Tomkins summarizes his paper as follows:

"Of particular interest are genetic relationships among species and early cultivars to determine if taxonomic classifications originally performed based on phenotype would be confirmed by molecular relationships obtained in the present study. Therefore, neighbor-joining analysis was carried out on the species and the early cultivars group. The resulting dendrogram is shown…. Taxonomy in the daylily has undergone recent changes and is still somewhat open to conjecture. For our purposes, the AFLP data will be discussed in the

context of recent classifications described by Erhardt (1992). Taxonomic classifications were generally supported by the AFLP data. The six H. fulva species all clustered together separately from the other species, which formed a separate cluster and were generally grouped according to Erhardt's proposed group classifications for the other species. Within this group fell the three early cultivars which showed close relationships to their respective ancestral species progenitors as described in the 1893 to 1957 Hemerocallis checklist... The only anomalies were two clonal variants of Hemerocallis citrina (var. Vespertina) and Hemerocallis dumortieri (var. Sieboldii)."

He then continues:

"Utilizing neighbor-joining analysis, the six H. fulvas were distinctly separated from the other species. Clustering within the fulvas also supported some fine-scale taxonomic classifications. For example, the distinction described by Erhardt between the two fulva double-flowered genotypes 'Kwanso' and 'Flore Pleno' is reflected in the molecular data. Within the middendorffii group, H. dumortierri, Hemerocallis middendorffii and Hemerocallis hakunensis all grouped together as proposed by Erhardt. However, the distinction between the citrina group and the middendorffii group was not well defined and contained some overlap. H. citrina and Hemerocallis minor were grouped together as proposed by Erhardt, but were also grouped with members of the middendorffii group. Erhardt had proposed a close relationship between two other members of the citrina group, Hemerocallis lilioasphodelus and Hemerocallis thunbergii, which was well supported by our data, but they did not closely group with the other citrina members. In fact, our data suggest that the middendorffii group and the citrina group should be merged into one large taxonomic group"

If one reads Tomkins carefully, there is a great deal of ambiguity present. He seems to be trying to keep with Erhardt but he continually diverges. Thus there are still open issues as to Classification using this data.

The purpose of this section was to summarize the work done on the genus Hemerocallis using genetic related probes in the process of determining the species and their interrelationships, namely using genes to study Hemerocallis systematics. We can reach several conclusions:

1. Use of gene related probes to assess the species in Hemerocallis has commenced. The use of AFLPs seems to be the most regarding at this stage.

2. There exists a multiple set of gene probes which permits the analysis of the genus in an exhaustive manner. Although RFLP and microsatellites and RSS are useful, the AFLP approach allows for massive screening. However mapping the genome is the ultimate

goal and then using microarray technology will ensure relationships can be studied in detail.

3. The use of a maximum likelihood approach provides most likely the best tool for assessing genetic heritage and in obtaining trees. This has its weaknesses but still is logically compelling and is as close to what we see in natural processes as well.

4. The use of microarrays and their derivatives will provide the best path to understand mechanisms of gene action between and amongst species.

5. Intra-species and intra-species variations are yet to be determined. Some of the studies focused upon show significant intra-species variation. This must be done in a more exhaustive manner to have better meaning.

4.8 Genetic Analysis of Species

We have developed a somewhat detailed analysis of the genetics of the plant and have also shown how the genetics and their paths can control the secondary pathway. We now pose an additional question or set of questions. Let us assume the following:

1. We have N species which have been agreed upon by the basis of phenotypic analysis. That is the mass of those who count have reached a consensus on what a species is and what plants are now considered a species.

2. We further assume that we have the complete genetic makeup of all of the species as well as a large base of material within each species. Thus we assume we have collected plants which we consider to be a single species and we have collections of collections which we consider to be species in their own right.

3. Let us assume we have further identified the gene segments which allow us to identify a plant in a species.

4. Let us further assume that there existed a single ancestor species from which all of the existing species arose by the normal process of evolutionary development. This assumption has not base in fact, it is merely an assumption that we will agree to until proven otherwise. We shall consider the counter to this assumption in a latter analysis.

Then we pose two questions.

1. If we agree to the above, then what were the genetic changes that would lead to a single originating species knowing the end point species as they are today?

2. Is the process which led to a solution to the first question, a unique answer or if it is not then what are the complete set of alternatives, if countable.

Then we can pose a third question:

3. If we know the evolutionary process backward, what is the set of possible evolutionary processes forward in time?

And then a corollary question:

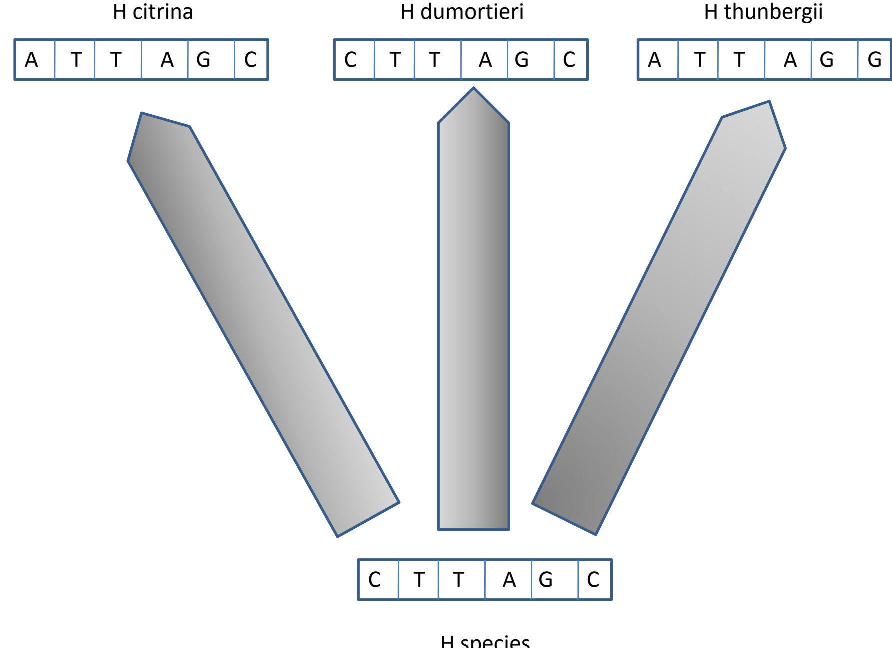
4. What impact does hybridization have on the forward evolutionary paths of the sets of species?

4.8.1 The Genetic Model

We now pose a simplistic genetic model for this problem. This is quite simplified but can be generalized with limited effort.

Consider four species and consider that the four species differ in three genes. We consider them species as shown below in genetic form. We show just a small part of a single gene. We also stipulate that it is this single gene that is controlling the plant's ability to adapt. This, of course, is highly simplistic.

We simplistically say that there are three species as indicated and that as a result of some genetic change they have mutated into the three species as indicated from a single species. We assume that we know what the single species from when it came is structured as.



Now we pose the hypothetical model for how this change occurred. Namely there is a probability that during each generation a mutation in a specific nucleotide may occur, a change from an A to T or the like. The model for this transition per generation is shown below.

4.8.2 Quasi-Speciation

Now we shall consider the concept of quasi-speciation. We will make the following assumptions:

1. We assume there existed a primal species and we assume we know the genome of that species.

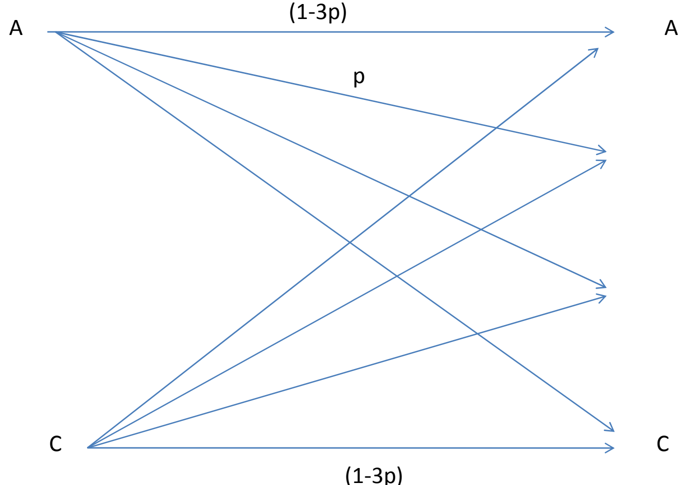
2. We assume that the genome controlling everything is the small segment we have shown earlier.

3. We assume that mutations occur at the rate p per generation. We also assume p is quite small.

4. We assume that the end states are the three species shown: H. citrina, H. dumortieri and H. thunbergii.

5. Now we assume that each of the terminal species is in a micro ecological environment. H. citrina is in an environment where the pollinating insects come out at night. In the day time there are many birds which devour the insects but the birds are diurnal and not nocturnal. In the case of H, dumortieri, we assume that the pollinating insects peak in early spring and that they have an ultraviolet sight which is attracted by the reddish brown on the petals of the flower. In the case of H. thunbergii we have insects which are diurnal but come forth in the mid summer and not too early. Thus certain characteristics are favored over others in each of the three environments. We must factor these into the model for the plant evolution, namely the element of fitness, selection fitness.

The Figure below depicts the genetic mutation model we shall use. It assumes that there is an equal probability for each nucleotide to change to another, and that there is a well defined probability that there is no change.



To those familiar with communications channels this is a modification of what is called the binary symmetric channel, BSC, a well studied artifact which is the basis of all modern telecommunications. We assume that the probability of a nucleotide changing into anything but itself is p and thus for the four possible end states we have three p probabilities and one 1-3p.

Now we define a term called the relative frequency of a specific gene combination, we call this:

$$x_k(t)$$

This is the percent or fraction of the population with gene configuration k. In our simple case there are 3 possible configurations.

We further stipulate that there is a term we call the fitness factor, which is a number we assign to a specific gene combination, relative to its ability to outgrow other genes.

The model used is that of Logistic Growth, proposed in 1836 by Verhulst[25]. If we use the same percent population as we described above then we have:

$$\frac{dx(t)}{dt} = rx(t)\left[1 - \frac{x(t)}{K}\right]$$

Where the constants can be chosen to satisfy the model. But simply, the model says that the percent of the population grows at some exponential rate until it gets too big and then it begins to decay at an even greater rate. There is no magic in this model just a statement of the facts which can emulate what we see in nature. The model is not based upon any underlying physical understanding; it is used primarily because it matches the curves.

Now let us introduce the concept of fitness. Remember our three species; one has night time insects, one has early season insects with vision fitness and the third has late season insects. Consider the first one, the nighttime insect environment. Let us begin by defining three elements:

Let:

$$x_1 = H \ citrina$$
$$x_2 = H \ thnubergii$$
$$x_3 = H \ dumortieri$$

These are the percents or fraction of the total population for each species. Also not we have assumed that the gene changes have already occurred and the problem we are focusing on is which one does best and thrives.

Thus we have:

[25] See Murray p. 2.

$$\sum_{k=1}^{3} x_k(t) = 1$$

Now we can look at the x values as probabilities as well, namely the probability one any one of the species at some time t.

Now let us assume that there are the three species and that any one of them may mutate into the other. Let:

$$Q_{i,j} = \text{Probability that } x_j \text{ changes to } x_i$$

Let us further assume that at time k we have a survival fitness which equals the probability that a species survives to time k.

Consider now two times, say k and k+1. Then we have:

$$x_1(k+1) = x_1(k) + Q_{1,1}S_1x_1(k) + Q_{1,2}S_2x_2(k) + Q_{1,3}S_3x_3(k) + \ldots$$

Where the S represents the respective survival probabilities. Finally we have a terminal factor to be specified. Now let us look at this equation from the perspective of incremental times, we obtain:

$$\frac{dx_1(t)}{dt} = \sum_{k=1}^{3} Q_{1,k} s_k x_k(t) + \ldots$$

Now let us focus on the final term. We know that the sum of the x's is zero. Thus if we assume that the last term is of the form of:

$$\frac{dx_1(t)}{dt} = \sum_{k=1}^{3} Q_{1,k} s_k x_k(t) + \left[\sum_{k=1}^{3} a_k x_k(t) \right] x_1(t)$$

Where the a's are to be determined. This equation is of the form we had presented earlier. Now we can sum up the equations by adding them together to obtain:

$$\sum_{k=1}^{3} \frac{dx_k(t)}{dt} = \frac{d}{dt} \sum_{k=1}^{3} x_k(t) = 0 = \sum_{k=1}^{3} \sum_{j=1}^{3} Q_{k,j} s_j x_j(t) + \left[\sum_{k=1}^{3} a_k x_k(t) \right] \left[\sum_{k=1}^{3} x_k(t) \right]$$

Now we can demonstrate that if the a's equal the -s's we have an equality since the Q's are a probability matrix and the rows and columns sum to unity. Thus we have the equation:

$$\frac{dx_i(t)}{dt} = \sum_{k=1}^{3} Q_{i,k} s_k x_k(t) - s^0 x_i(t)$$

$$where$$

$$s^0 = \sum_{k=1}^{3} s_k x_k(t)$$

This is called the Eigen-Schuster Equation or the Quasi-species Equation[26].

Now let us consider a simple example.

Example:

We assume we start with three species in equal amounts. Somehow the mutations that led to these three were from some parent and that what resulted were three differing species each with a gene that allowed for a specific blooming time. The H. citrina at night, the H. dumortieri in early summer or late spring and the H. thunbergii in midsummer in the day time. We then send these three species into an environment which has summer mid day insects.

Let us defines the survival statistics as follows:

$$s_{thunbergii} = P\left[pollination \text{ H. thunbergerii / Region}_N\right]$$
$$s_{citrina} = P\left[pollination \text{ H. citrina / Region}_N\right]$$
$$s_{dumortieri} = P\left[pollination \text{ H. dumortieri / Region}_N\right]$$

We could now perform an experiment in the field to determine these values. We would watch the insect population, see how they are attracted to the plants, and perform a count on the number of insects setting on any plant. From this analysis we could then determine the above survival values for each species. In this example we have pollinating insects peaking in mid-summer consistent with a maximum of the s for thunbergii and minimizing the opportunity for the other two.

First, the transition probabilities are given in the matrix below. The probability of transitioning is quite low.

[26] See Eriksson et al. This paper is an excellent summary and extension. In contrast, the book by Nowak is filled with so many errors that as an introduction it is both confusing and a detriment to learning.

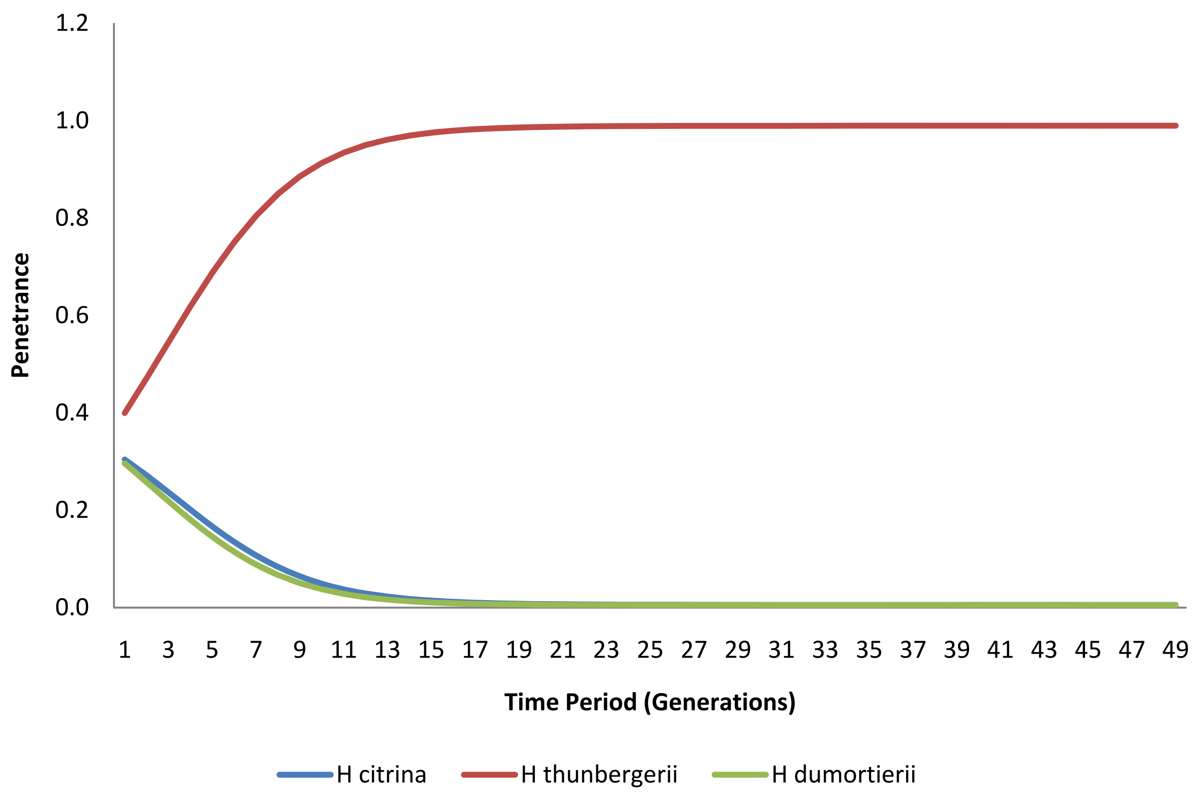
	Q1	Q2	Q3
Q1	0.9900	0.0050	0.0050
Q2	0.0050	0.9900	0.0050
Q3	0.0050	0.0050	0.9900

These above measures are assumed determined from measured mutation rates of the gene in question.

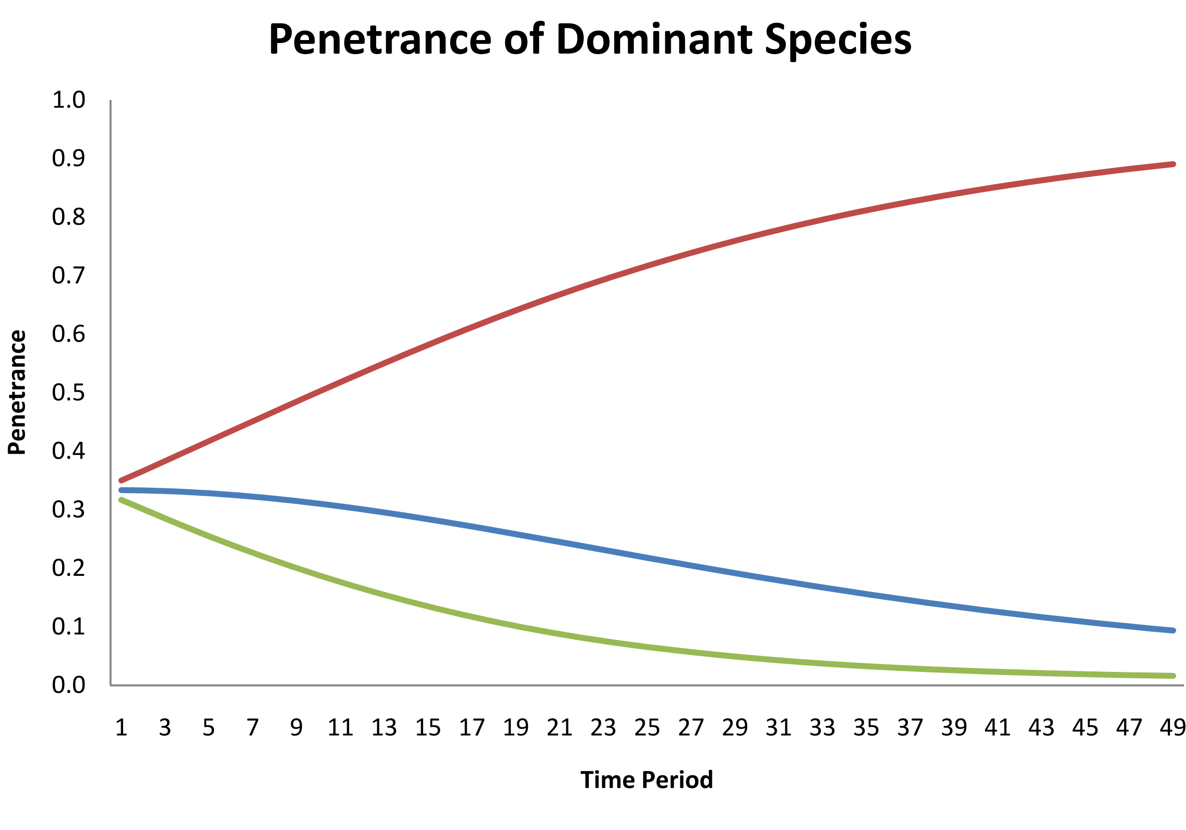
Then the fitness measures are presented below. These are determined from field work as we have discussed previously.

S1 H. thunbergii	S2 H. citrina	S3 H. dumortieri
0.3220	0.0310	0.0070

We have assumed we obtain the following characteristics for the penetrance of the species over time. In just 15 generations we have almost 100% penetrance of H. thunbergii.



This Figure shows how the dominance of one species occurs if there is a fitness. Consider now the same case but with a difference set of fitness factors. The results are below:



Here we have used the following fitness factors:

S1 H. thunbergii	S2 H. citrina	S3 H dumortieri
0.2000	0.1500	0.1000

H. thunbergii still ends in the dominant position but not as fast. In this case we have assumed that each of the species has a modest fitness. Convergence always occurs to the dominant species as long as one is dominant.

Now if we look at this simple example we see the dynamics of one species taking over another set of species due to a single identifiable and quantifiable factor. One suspects that this simple model is not the way nature functions. In fact there may be an agglomeration of many such factors, some enhancing and some detracting from the growth of the population.

4.8.3 GENETIC CHANGE

Now let us consider the issue of genetic change. We can look at this problem as follows:

1. We assume some initial state of a gene and then we look at the possible end states at some latter time.

2. We see several end state genes and we ask what the original state was.

In the first question we have a step wise stochastic process with some collection of end states. Unless we apply some form of fitness we will never get a convergence on a finite set to occur. In addition, the fitness function may be changing with time. In the second case we have end states and we than must try to obtain a conversion which takes us back to the same single starting point. This, however, assume that we have but a single start. There may actually have been multiple divergent starts.

For the second question let us analyze a simple example. Let us assume a single yet unknown initial state. For our analysis let us assume that there are N possible states, and that we indicate each of them by a number i. Let us further stipulate that the percent of any one state is a function x which is akin to what we have defined in the Eigen Schuster equation.

For example, let us assume that the gene is K nucleotides long and there are thus 4^K possible states. Thus the state vector is given by:

$$x(k) = \begin{bmatrix} x_1(t) \\ \ldots \\ x_N(t) \end{bmatrix}$$

$$N = 4^K$$

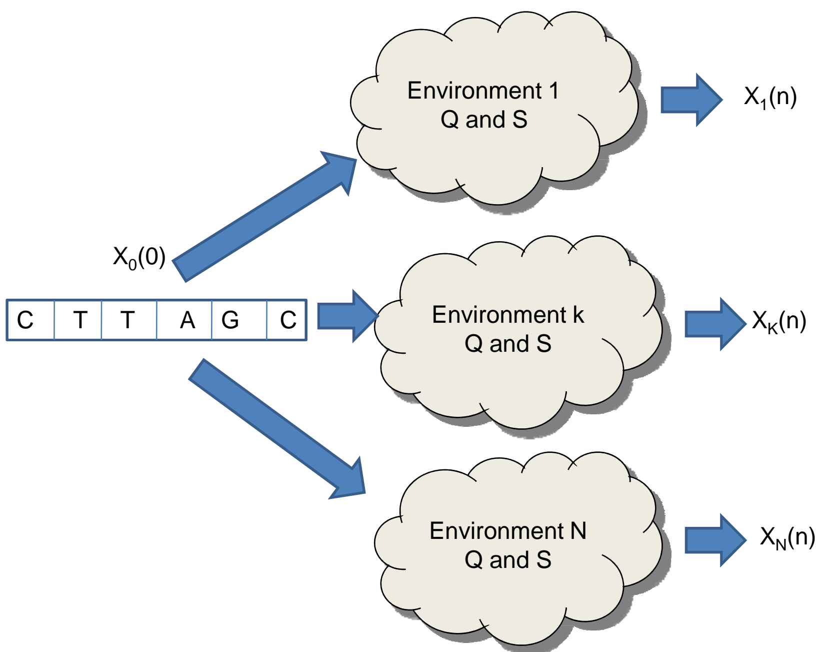
Using the Eigen Schuster equation we obtain:

$$\begin{aligned} &x(k+1) = x(k) + A(k)x(k) - s_0 x(k) \\ &where \\ &A(k) = Q(k)S(k) \end{aligned}$$

Or we can write it as:

$$x(k+1) = B(k)x(k)$$

Now we assume M environments. Each having a different set of s elements.
The problem is as we describe below. Assume there is a starting sequence. We then have the starting sequence go through N separate environments as shown below. Each environment is controlled by an Eigen Schuster equation. The net result is N stable species surviving uniquely in each environment. It should be noted that in an Eigen Schuster environment a single surviving species is always the case if that species has a dominant fitness. That assumption we are using here.



Before continuing let us consider the model a bit further and then consider the mutation process as well. For the model we assume a K long gene segment having 4^K possible alterations. We can simplify this with a binary scheme with 2^{2K} possible outcomes. Thus we can map:

$$G \to 00$$
$$A \to 01$$
$$T \to 10$$
$$C \to 11$$

Then we can assume, with no loss of generality that the initial state is:

$[0000000000..00]$

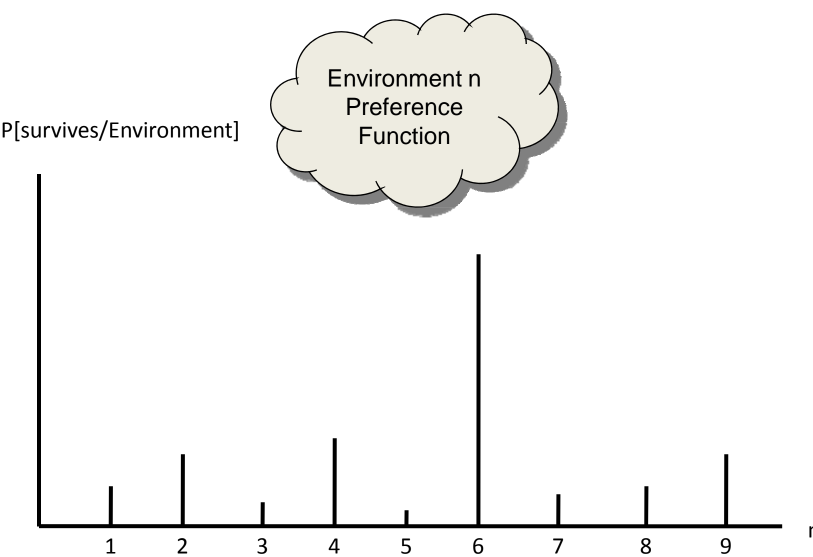
We now will ask how this relates to the Eigen Schuster equation, for in that case we are focusing on changes back to itself.

Let us define states for each environment. That is we define:

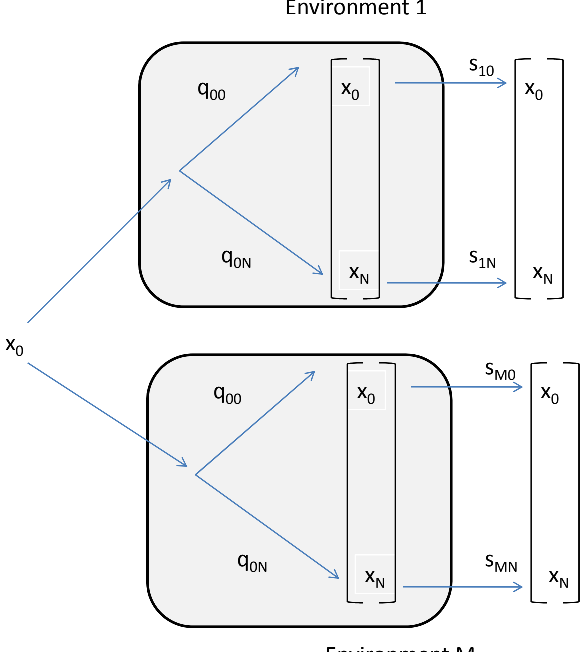
$x^i_j(k)$= the ith environment and j^{th} state at time k.

We assume we know the end states. Now let us look at environment m and its end state j, specifically $x^m_j(k)$.

Consider now the Figure below where we have 9 possible genes each with a fitness function defined. The gene 6 has the largest and we then know that given the Eigen model we would converge on that gene dominating the total population.



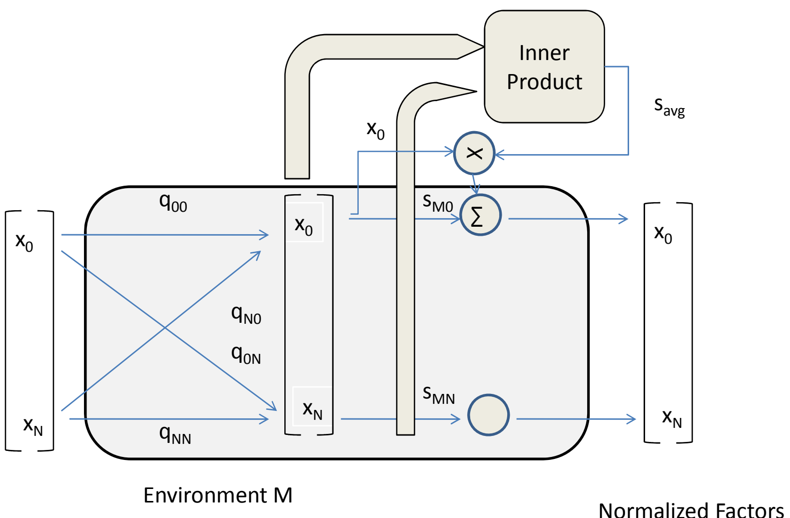
Now we can take the Eigen model a bit further. Take M environments and take a single step. Then we have M possible transitions each with its environmental fitness functions as shown below.



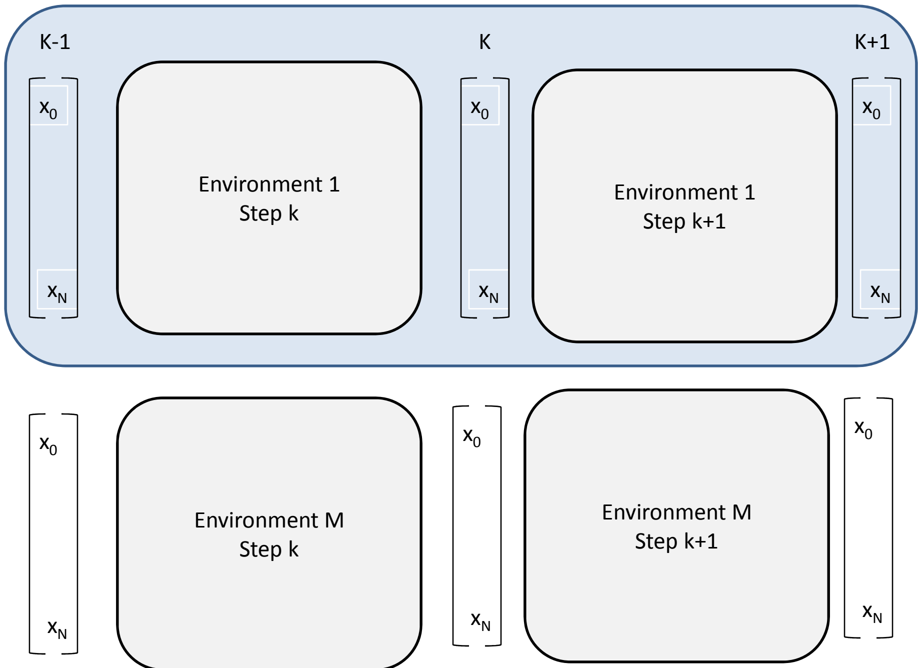
Then we can use the Eigen model to obtain what we see in the following. The transition probabilities are multiplies by the s values for each gene in each environment and then we must subtract the normalization factors. After a period of time we then obtain stable species in each separate environment.

4.9 Conclusions

In this Chapter we addressed three issues. First we examined the tools available to determine genetic metrics which can be subsequently in classification. These tools allow us to look at the gene level in various species and based upon that have a solid repeatable means to ascertain the relationship between species. Then we examined the various classification methodologies. There are two extremes here. On the one hand we can use the genetic generated markers and apply them to classification methodologies using some form of closeness algorithms. Second, we can attempt to generate the actual evolutionary sequences that led to the end points we have from the markers and attempt to ascertain what the best for the predecessor is. The latter case makes certain assumptions which we may not validly agree to at this time knowing how nature actually functions. Finally we look at some analytical models for gene pools surviving. This is an alternative look at the process of what happens if one gene is favored over another. This final discussion was engendered when I tried to pack too many species in a single display garden and saw that H hakuensis just pushed out H altissima and just dominated any H minor.



The step model takes the previous diagram and then step by step creates what we have shown below.



Now we can take any specific initial species forward in time. Unlike the prior example where we started with a mix and then allowed both change and fitness to effect a change, here we start with a single species and allow both to migrate the species over multiple environments. There is not initial environmental stable species here. In fact we may assume that the initial species in environmentally unstable.

5 FLOWER COLOR AND ANTHOCYANINS

This Chapter looks at the issue of the coloration of plants and the ability to estimate the concentrations of certain colorants such as anthocyanins based upon commonly available spectrometer methods. The approach is to begin with classic color theory which has been employed extensively elsewhere and then to develop a model for reflectance using the Beer's model and in turn provides a set of methodologies to estimate the concentrations of all colorants in a cell. This approach can then be employed in several areas; first in the determination of the genetic networks generating the colorants, the gene expression identification problem, and secondly the issue of flower color patterning, namely tessellation.

5.1 INTRODUCTION

Flower color is a direct result of light absorption in the cells of the petals and sepals. The process of absorption may be complex and many but the result is that the reflected light spectra will absorb certain parts of the visible spectrum and allow other parts to be reflected back out by the cell walls. To some degree there are many complex and yet to be understood or characterized processes at play. However, if one is seeking to estimate the concentrations of the chemical elements which lead to the coloration, then it is possible to do so using the means and methods proposed herein.

Once the concentrations of such elements such as the anthocyanins has been determined then it is possible to use that data and work backwards to assess and determine the nature and workings of the genetic pathways which have given rise to these colorants (see McGarty, 2007). In this paper we develop a method to determine the concentrations of colorants resulting from secondary pathways in flowers. The method employs the use of standard spectroscopic techniques and using the basic principles of color absorption provides a detailed set of methodologies to estimate separate concentrations. It is assumed that for each secondary pathway colorant its individual profile of extinction or absorption is known.

In this paper we address several issues. The objective of addressing these issues will be twofold. First we will need the understanding to proceed to the issues of understanding genetic pathways and to understand and explain the complex issue of flower color. The second use will be the establishment of a base to perform measurements and analyze the resulting data to validate the theories. Thus the issues we must join are:

1. **Color and Human Factors:** In previous work we used the Tristimulus model to analyze the results obtained from the measurements in the genetic pathway efforts. We argue here that this is a limited approach which on the one hand must be

understood and integrated in what we are doing but on the other hand must be gone beyond if we hope to obtain the resolution required.

2. **Color in Plants:** This discussion is a complex set of issue regarding plant color. On the one hand we review and position the anthocyanins and other colorants and on the other hand develop constructs for explaining the passage, absorption and reflectance of light as color in plants.

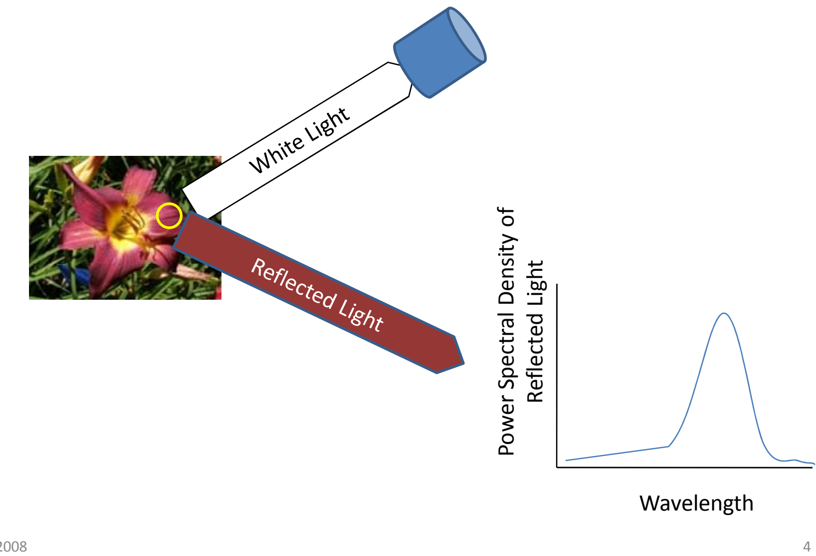
3. **Measurements and Methods:** Spectrometry has been a mainstay of assessing molecular structure especially of complex organic molecules. We review the physics of the underlying phenomena and then review the experimental techniques employed. We argue that the use of Fourier Transform Spectroscopy is best suited in this environment.

4. **Data Analysis and Concentration Estimation:** Once the data has been captured, we then seek methods using the inverse of Beer's law to ascertain the concentrations of and types of anthocyanins and other colorants in the cells. Multiple methods are presented, developed and compared.

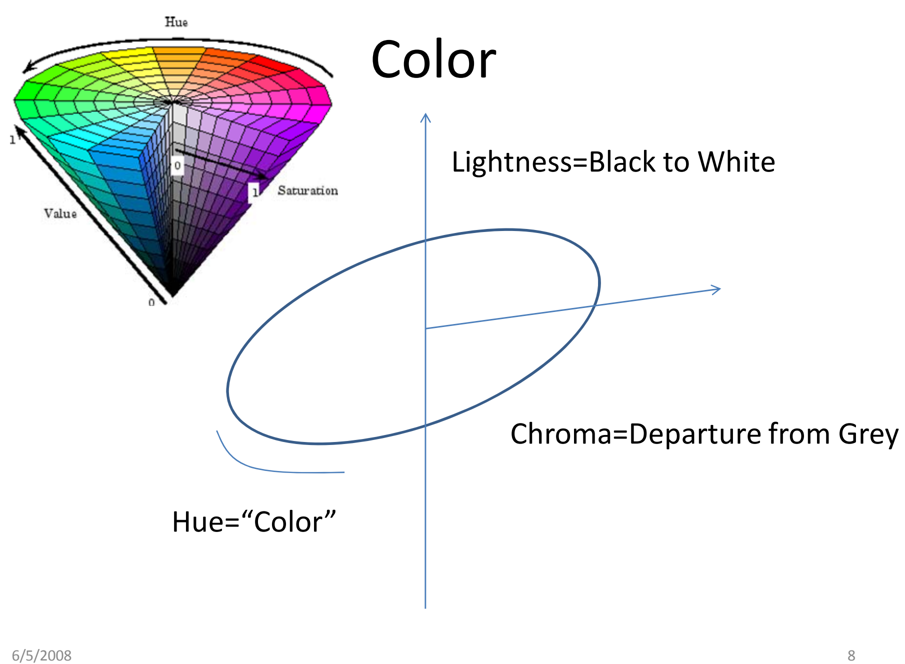
5.2 Classic Color Theory

Color can be viewed from several perspectives and the two focused on herein are the human eye and the measure power spectrum. The human eye views color in a complex manner since the eye receives color stimuli via sets of sensors which are tuned to three possible visible frequencies, the classic red, green and blue.

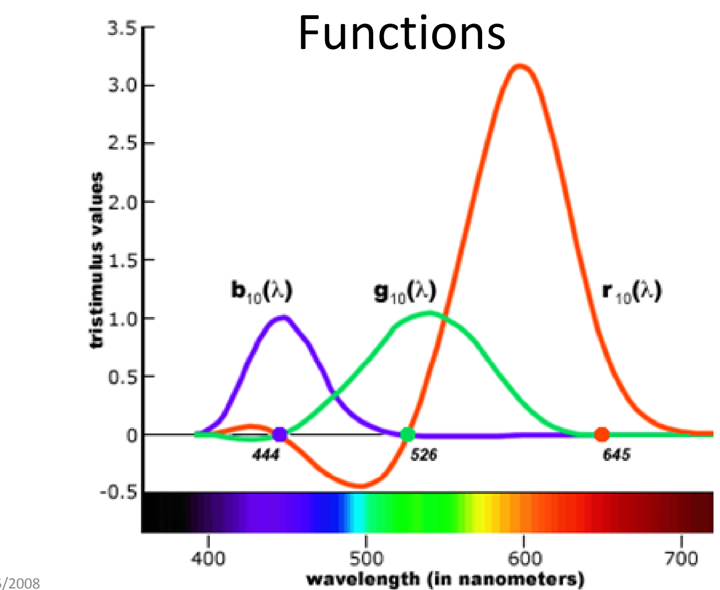
The Figure below depicts the basic concept. We see a flower as a certain color. There is "white" light shown upon the flower and the light is reflected from the petals and sepals and what we perceive is a red flower. This perception is a combination of two things; what part of the incident white light is reflected, and how our eyes process that reflected light. Thus color has two meanings for us. First color is nothing more than a reflected spectrum of electromagnetic waves in the optical frequency band. Second it is what we perceive as a human observer and in turn name as a color. The latter approach can and often is quite subjective. This latter approach is the basis of print color, paints, dyes, pigments and the like where the end point is the presentation of some artifact of a desired color. It is the former or first approach we seek to use, namely what is there independent of the observer, specifically the spectrum.



In the context of color as perception there is a collection of terms which should be understood. Hue is a synonym for a color. Red is thus a hue as is orange. The hues cover the visible spectrum. Then there is the lightness, ranging from whit to black. The third element is the chroma which is the departure of the hue from gray. We show these elements below.



The color school then takes three more steps. These are the CIE models for color. The first step is the Tristimulus models. The Tristimulus function is shown below. This is NOT a spectrum. In addition negative values mean more positive stimulus. These are also the result of extensive experimental modeling. The red, green and blue Tristimulus model as shown below characterizes three stimuli which affect the three receptors of color in the eye.

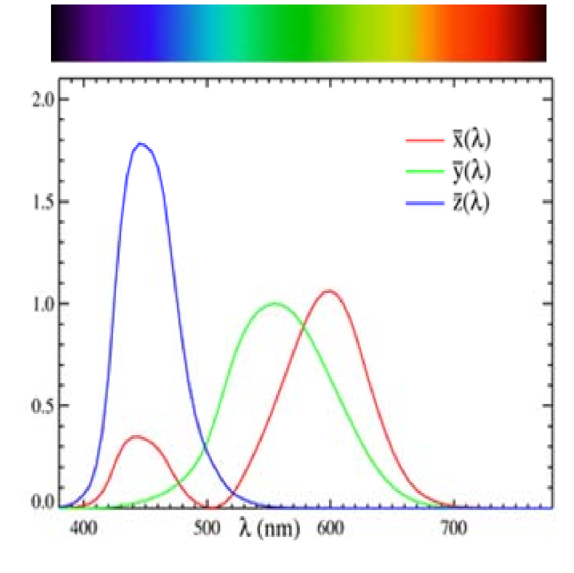


The above Tristimulus curve, called the color matching function, were experimentally determined through an experiment where a person looked at a test color centered at a wavelength as shown on the horizontal axis above and tried to match it by adjusting a red, blue and green lamp in the reference field. This could be accomplished for all regions except between 444 and 526 nm. In that region a red light had to be added to the test field to adjust the color to match. In effect the test color was changed by adding red. This adding of red is accounted for by the negative portion of the above curve[27].

Following this above model based upon experiment is the spectral approach called the standard observer consisting of the X Y Z model as shown below. They do effectively represent quasi spectral responses since they are all positive. It is possible to transform between the RGB and the XYZ formats.

[27] See Berns p. 49.

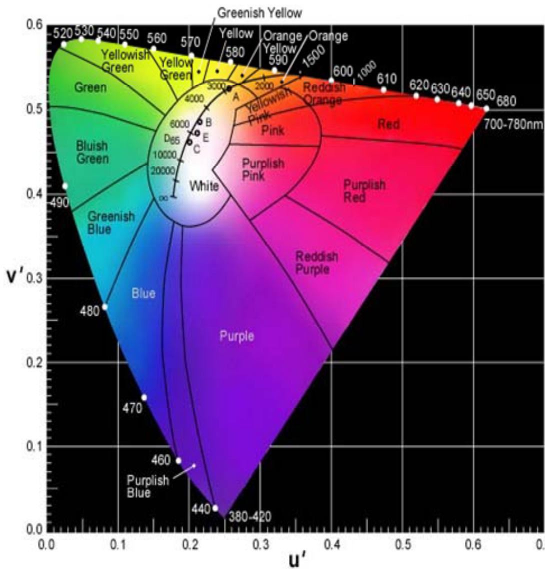
CIE Standard Observer Curves



6/7/2008 12

The CIE Chart is a manifestation of how the three stimuli above can be added together to create a broad set of colors, hues. It is possible to go from red, thru green and then to blue. One need only mix the three stimuli in the proper ratio. Then other hues can also be generated.

CIE Chart



6/5/2008 17

To understand this a bit better we analyze the RGB system first. We start with a source specified by intensity I dependent on wavelength. This allows us to define the following:

$I(\lambda) = $ *the spectrum of a specific sample*

Define

$$R = \int I(\lambda) r(\lambda) d\lambda$$
$$G = \int I(\lambda) \overline{g}(\lambda) d\lambda$$
$$B = \int I(\lambda) b(\lambda) d\lambda$$

Note, as we had stated, the color matching primaries show negative values because the negative was the way the CIE arbitrarily represented an excess positive contribution required to be added to a primary to achieve the desired spectrum response while keeping the elements normalized. Specifically:

$$\int \overline{r}(\lambda) d\lambda = \int \overline{g}(\lambda) d\lambda = \int \overline{b}(\lambda) d\lambda$$

Now this also implies the following are to be true:

$$r = \frac{R}{R+G+B}$$
$$g = \frac{G}{R+G+B}$$
$$b = \frac{B}{R+G+B}$$
and
$$r + g + b = 1$$

In a similar manner we can do the same for the XYZ system. This is done as follows:

$I(\lambda) = $ *the spectrum of a specific sample*

Define

$$X = \int I(\lambda) \overline{x}(\lambda) d\lambda$$
$$Y = \int I(\lambda) \overline{y}(\lambda) d\lambda$$
$$Z = \int I(\lambda) \overline{z}(\lambda) d\lambda$$

And as was the case for RGB we also have the normalizing factor. Not that it is this normalizing factor which assures our ability to deal with the triangular plot of color.

$$x = \frac{X}{X+Y+Z}$$
$$y = \frac{Y}{X+Y+Z}$$
$$z = \frac{Z}{X+Y+Z}$$
$$and$$
$$x + y + z = 1$$

Finally there exists a set of transforms which allows one to convert from one to the other. This is shown below:

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = A \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

A such that sums of r,g,b and x,y,z are unitary

Therefore for any color C we can write it in one of the following two manners:

$$C = rR + bB + gG$$

$$C = xX + yY + zZ$$

Thus an x,y plane can be constructed such that any color can be characterized as a pair of coordinates (x,y). This is the CIE Chart which we have shown above. It must be noted that all of this analysis is predicated on how "we" see color and not in any context of how it is created or the underlying physics of color.

Now there are two other brief examples worth noting. First is the concept of additive colors, such as those we see when we add lights. This was the basis of what Newton did in his early experiments. By adding lights we can ultimately create white. We show that below.

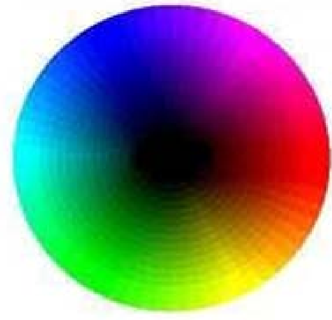
Additive Primary Colors

Additive Colors combine to form white. Traditionally adding lights is additive whereas adding colorants, pigments or dyes, is subtractive.

6/5/2008 19

The opposite is the subtraction of light, and this is the result of adding pigments of different colors together in an oil painting. If we were to add all the colors together then we obtain black and not white. This is subtractive, for we are in reality removing colors by the use of those pigments. In many ways this is the difference between water colors and oil paint. We show this subtractive result below:

Subtractive Primary Colors



Subtractive Colors form black. Subtractive mixing involves the removal, subtraction, of light from the mix. Removing all light ultimately results in black. Absorption only is called simple subtractive mixing whereas combining this with scattering is complex subtractive mixing.

6/5/2008 20

Notwithstanding the above detail and its use in many industrial processes, these methods used in classic colorimetry are methods that rely upon the human by necessity being part of the process. We when looking at plants, shall disregard the human.

5.3 Spectra and Measurements

The measurement of absorption spectra can be accomplished by a variety of means. We present here two methods; classic spectrometry and Fourier Transform Spectrometry (FTS) also called Fourier Transform Infrared Spectrometry, FTIS. However FTS can be applied to the optical bands as well.

The goals using these methods are as follows. First to determine the absorbance and extinction coefficients of the secondary products that are colorants. This means that solutions of purified anthocyanins, Peonidin for example, would be used and their absorbance and extinction coefficients determines for all wavelengths over the optical band[28]. This is accomplished for all targeted absorbents. Second, perform the same on all known colorants found in a target plant cell. This could include any secondary product or even proteins which have absorbent properties. Generally the other chemical elements react in an absorbent manner out of the optical band. Third, perform the analysis on target cells. Our approach is to perform this on a cell by cell basis thus requiring focused optical positioning.

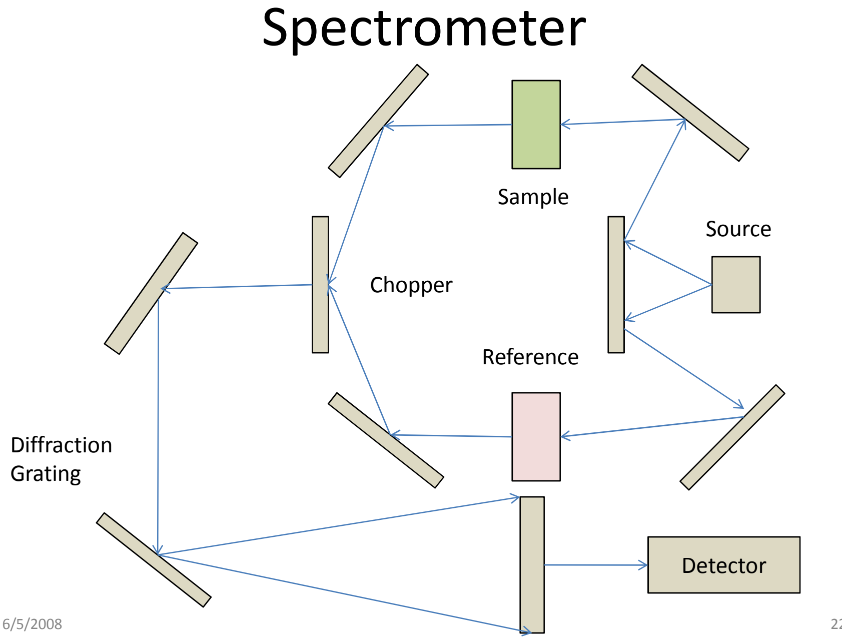
5.3.1 Classic Spectrometer

We first look at the classic spectrometer. It uses two paths for transmission, one through a cell with the target secondary and another cell without any secondaries or colorants. The second cell is a reference cell. The reason for this approach is to calculate the difference in absorption. The spectrometer is shown below. It functions as follows:

- Select a Reference and a Sample. Source Spectrum is to be determined using reference.
- Send light from source through both sample and reference. The source must be broadband wavelength. There will be no need to regulate amplitude across the band since a difference signal is obtained and the result will be expressed as a ratio.
- Chop the signals using electronic chopper so that half interval it is sample and other half it is reference. This can be accomplished with a time controlled electronic device or even a mechanical rotating wheel which can be synchronized to the measurement elements.

[28] See Cantor and Schimmel, Part II, pp 380-388.

- Send to Diffraction Grating to spread out signals over visible spectrum.
- Sample from one end of spectrum to the other by mechanically sampling the diffraction grating spread out. Remember that the diffraction grating act as a prism and spreads out the signal spatially over the optical band.
- Use the reference as the baseline and then measure the ratio or the difference of sample to reference and plot. This generally requires just a difference amplifier at the measuring point and synchronizing it with the chopping signal.



The spectrometer, as shown above, functions well for the determination of relative absorption. It is a long and sometimes cumbersome process because the screen in front of the detector is scanned slowly and this provides the signal used to ascertain the difference measurement. There is an issue of accuracy and precision in the collection of data and there is also the issue regarding the amount of light intensity requires. One should remember that as we spread the light out through the grating we see the spectrum now spatially but in so doing reduce the signal strength of each segment. The spectrometer has advantage and disadvantage in this configuration.

5.3.2 Fourier Transform Spectrometer

The FTS is a more recent embodiment of a spectrometer and it eliminates many of the accuracy and precision issue of the classic spectrometer. In many ways it may be viewed as a mini-CAT scanner in that it collects data which is the Fourier Transform of the desire waveform, namely the absorption spectrum.

The FTS works as follows:

1. A target sample is placed in front of a detector. The detector is a broadband detector and it provides at its output the integral of all the power entering across the optical spectrum. The optical spectrum will be the target spectrum of interest so we delimit the detector to that. We also assume we know the detector response and that this can be adjusted for by means of signal pre-emphasis. This means that the detector works as follows:

$$P = \int \tilde{S}(f)df$$

Here P is the total power and S(f) is the power spectral density of the combined signal. We will look at that in some detail in a moment. Now we assume that the detector may itself have a spectral sensitivity given by H(f). Thus what we really receive if we do not pre-process is:

$$P = \int \tilde{S}(f)H(f)df$$

Which may bias out answer? The way to avoid this is to do some pre-emphasis on the front end by using filters which do the following:

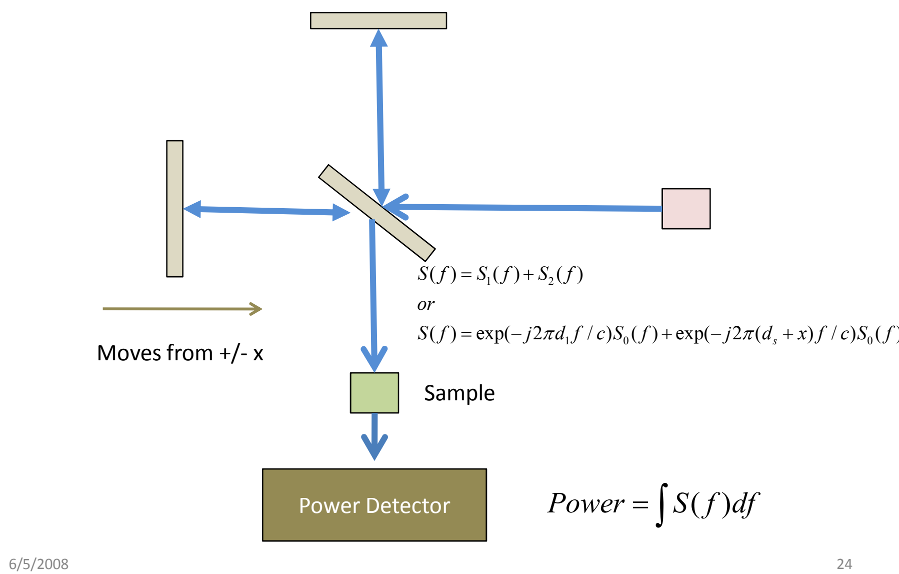
$$P = \int H(f)G(f)P(f)df$$
$$= \int P(f)df$$
$$if$$
$$H(f)G(f) = 1$$

This is optical pre-emphasis filtering as one does with FM radio. This is a standard approach.

2. Now let us go back to the input. We assume we have a flat frequency broadband source of radiation emitted from the source. If now we can also pre-emphasize that as well. Then this source follows two paths. Path 1 is a fixed path up and down and through the sample. Path 2 is one that goes to a reflector whose portion is changing uniformly in time and is accurately measures. This second path then send the same signal with the sole exception that it is phase offset from the main path. At times it may be totally in phase and at time totally out of phase. For every position x of the reflecting mirror we measure the combined power spectrum received, the integral of both signals, measure as their amplitude.

3. It can be shown, we do so below, that if one collects the P values and notes them as P(x) then if we sample x properly we obtain samples of the Fourier Transform of the S(f) function. Thus collecting P(x) for the correct values of x and doing so with enough

samples we can then perform an inverse Fourier Transform to readily obtain S(f). This is FTS.



The details can be displayed simply as follows. The signal received is the direct and the reflected and they are combined in complex space to account for the phase difference as shown below:

$$S(f) = \exp(-j2\pi d_1 f/c)S_0(f) + \exp(-j2\pi(d_s + x)f/c)S_0(f)$$

$$\approx S_0(f)\cos(2\pi fx/c),\ \textit{so that the received power is:}$$

$$P(x) = \int S_0(f)\cos(\frac{2\pi}{c}fx)df$$

The result is that P(f) is the real FT of S(f).

FTS shows that we can obtain P(x) and it is the Fourier Transform of S(f) the absorption spectrum of the sample. We take P(x) for many values of x and then inverse FT.

5.4 Plants, Color and Chemistry

We can now consider plants and their colors. We have discussed this before and it is covered generally in the literature. However we want to focus on specific colorants.

5.4.1 MOLECULAR ISSUES

We begin with the most common colorant, chlorophyll, the element which makes leaves green

Chlorophyll

Chlorophyll contains many bonds and many outer shell electrons which in turn can absorb many photons across the incoming spectrum of visible light. They are absorbed by causing electrons to jump to more distant orbits and by exciting bonds between the constituent atoms. The result is the absorption of photons selectively across the optical spectrum.

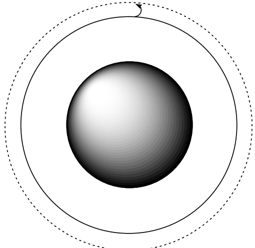
6/5/2008 27

What causes absorption? The answer is a complex one but it may be simplifies into two parts/ first is the excitation of outer layer electrons, using the exact energy of the incoming photon. The second is the resonance excitation of the bonds in the carbon elements especially. There is a complex set of issues here. The correct manner to approach this is via the quantum mechanical methods. They are extremely complex due to the complex structure of the compounds. However phenomenologically it is easy to measure.

We show these phenomena in the Figure below.

Absorption

- Molecular: Electrons and bonds
  - Electrons in outer shells



$$\varepsilon = h\nu = \frac{hc}{\lambda} = energy\ of\ photon\ absorbed$$

  - Bond vibrations

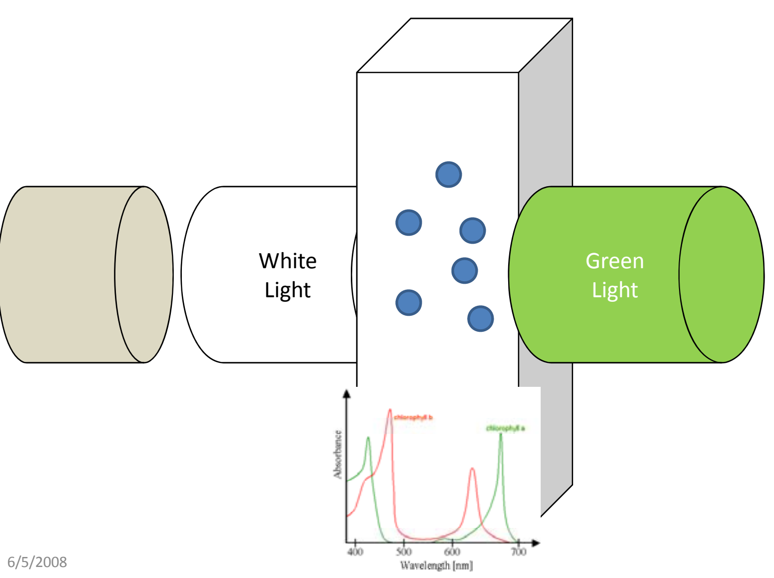
6/5/2008 28

5.4.2 Absorbance and Reflectance

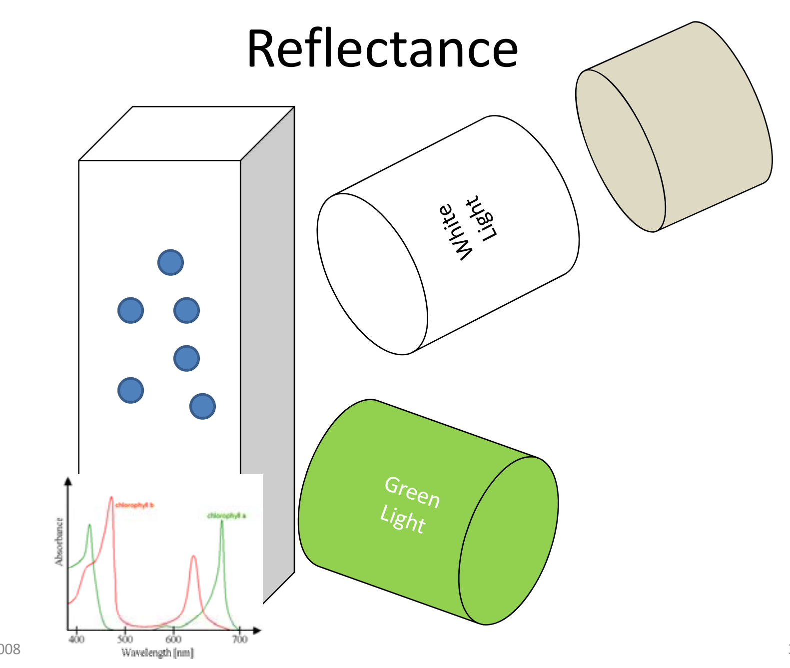
The next question is what are leaves green and flowers red? The answer is simple obtained by understanding absorbance and reflectance. In the figure below we show white light passing through a cell or sell filled with chlorophyll. Chlorophyll absorbs red and blue light and lets the green part of the spectrum pass unabsorbed. The absorption results from the very atomic interactions of the photons on the structure of the chlorophyll molecule we have shown previously.

In the Figure below we show a white light as we may see in the sun and we shine it upon a cell or set of cells and we assume that it is absorbed but manages to pass through. The light emitted is green.

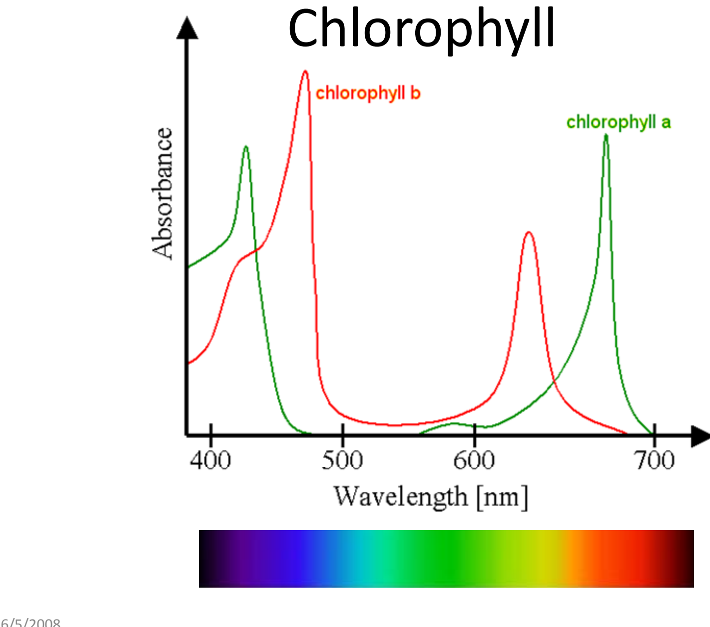
Absorbance



In contrast to absorbance we also have reflectance. Reflectance results when we shine white light upon a cell or set of cells and the light is refracted and in turn reflected back out into the general direction of the incident light. In the Figure below we show white light impinging on a cell and the cell is filled with chlorophyll. The chlorophyll absorbs the red and blue so the reflected light is green. This has the same characteristics as Transmittance.



The absorbance of chlorophyll is shown in the following Figure. We shall define absorbance latter. However what we see is that chlorophyll absorbs the low and high frequencies, the long and short wavelengths, and leaves the middle wave lengths relatively un-absorbed. This means that what is reflected back from a plant cell composed of chlorophyll is primarily green. Thus the green leaves of plants. Of course we will not see the same in flowers. In addition, in the fall when the chlorophyll degrades as the plant goes dormant, what is left in the leaf are the anthocyanins which we shall show have a reddish orange tint.



The above Figure also depicts the two types of chlorophyll, but for our purposes we need not be concerned with them.

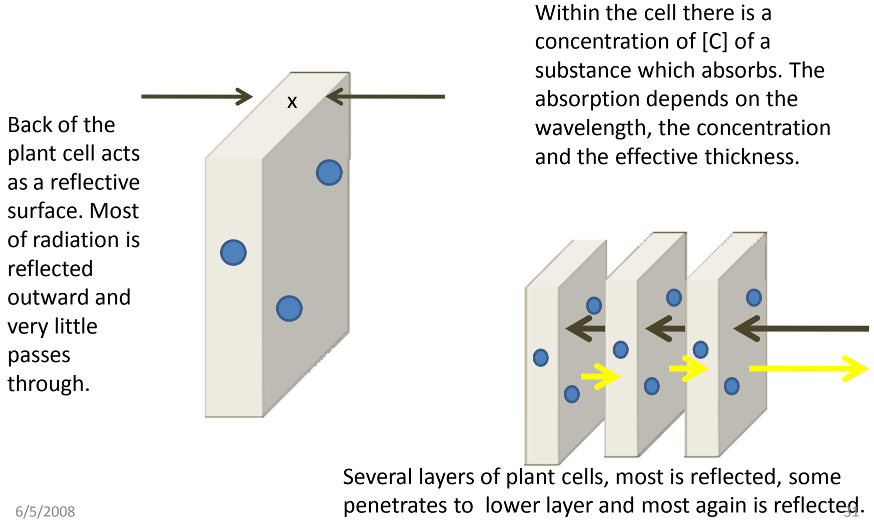
5.5 Plant Cell Reflection

The next issue is to understand how reflection occurs in plants. This is somewhat of a complex issue but has a simple explanation.

A plant cell on the surface of a plant has a cell coating composed of cellulose, hemicelluloses, pectin and proteins which are all relatively transparent to light in the visible spectrum[29]. Cellulose is a long chain of glucose residues which form a ribbon.

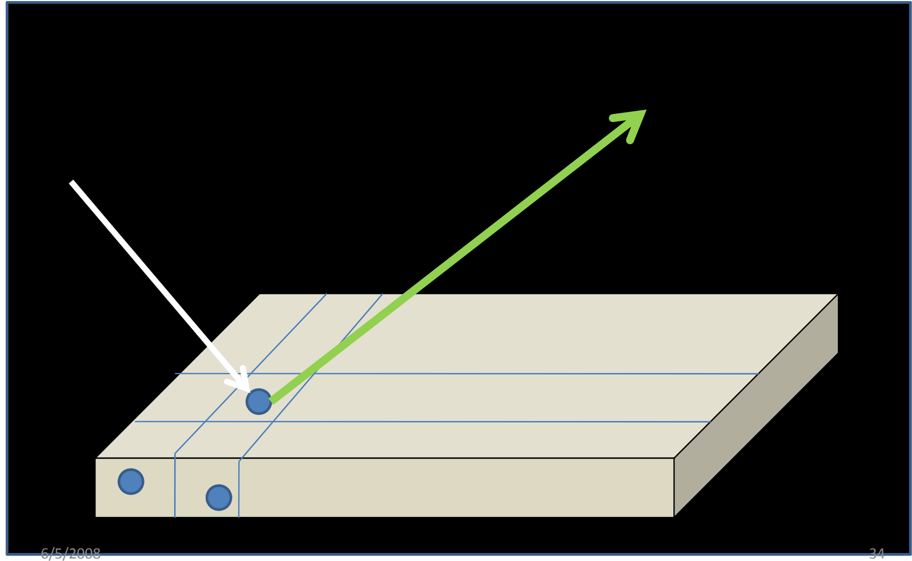
[29] See Taiz p. 22.

Visible light penetrates this wall easily and then is passed through the cell. The light then is reflected back out, most of it, and some continues to flow to lower layer cells.



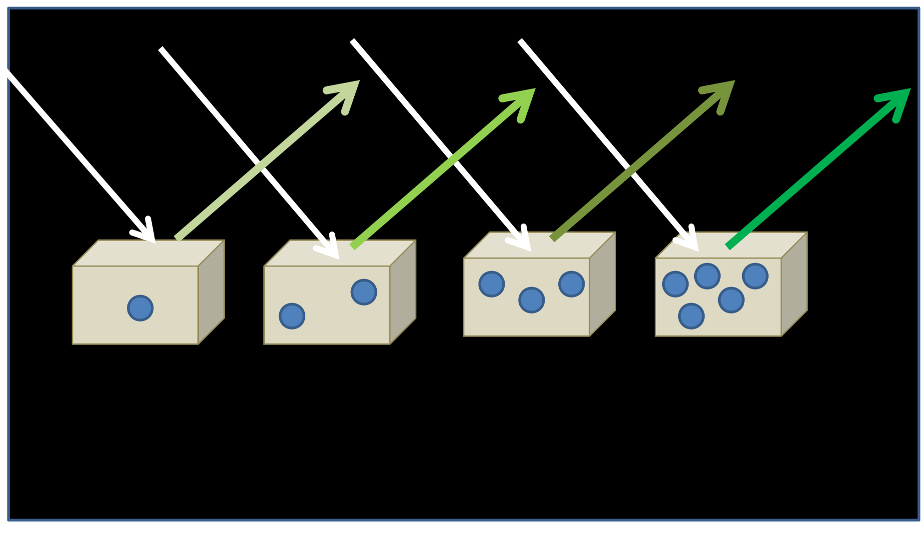
Reflection occurs in a plant cell as a result of the standard process of light being reflected at the boundary where we have an abrupt change in the index of refraction. This index changes as we go from air into the cell, and indeed even in the cell itself. The cell has water, proteins, colorants, a nucleus and many other constituents. The refraction at the interface changes the direction of propagation and the change may reflect the light out or further into the cell. The issue is dependent upon the incident angle and the index of refraction. We show a prototypical example below.

Cells



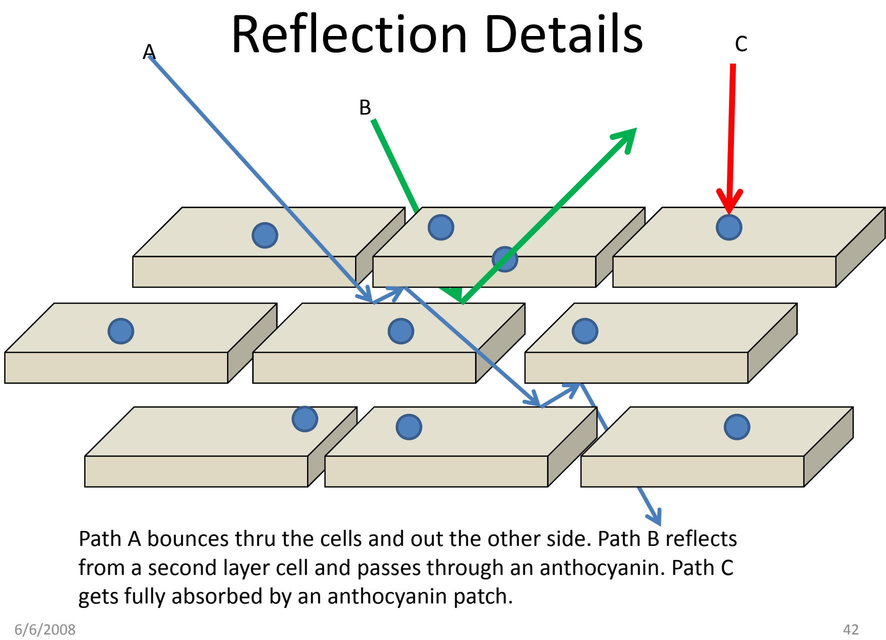
The Figure below shows another phenomenon. As the cell concentration increase more absorption occur and the color may actually change. Thus cell concentration is a major factor. In addition the thickness of the cells and the number of total cells will also be a significant factor and we shall discuss that next.

Cell Concentration



6/5/2008 35

The process of intracell reflection is somewhat complex to say the least. It has been studied for over a century and there are still many theories to explain various cell structures. However for the case of a flower as Hemerocallis the presentation can be simplified. The Figure below, adapted from Lee, shows three paths through several layers of cells. One path bounces about and finally is reflected albeit attenuated by the colorant molecules. A second path manages to go through the cells and out the other side,. and a third path gets fully absorbed.



The details of the above are generally difficult to analyze due to the random nature of the cells and the colorants. McGarty (1971)[30] provides a summary of the approaches. However there are certain metrics which can be useful. The plant cell has a cellulose wall which is rigid. The dominant substance inside the cell, especially in a flower, is the vacuoles filled with water. The colorants are mixed in the cell. A plant cell is about 100 µm in width and about 20-50 µm in depth. The wavelength of light in the visible region is 500 nm which is 20 time less. Thus scatter is not a major factor. However refraction and reflection are. The index of refraction of the cell is about 1.2-1.4 (see Lee)[31] and from Born and Wolfe (p. 41) we know that for an interface of this type the reflection and refracted components can be calculated. If we define A as the incident amplitude,

---

[30] See McGarty, 1971. This Thesis details the complex issues of multiple scattering in complex media such as a cell matrix. The Thesis also summarizes the experimental and theoretical work to the date of publication. Some extensions have been made since that time but the solution to the problem is still somewhat intractable except in a statistical sense. We use Beer's Law in the next section as a means to handle the complex nature of the optical problem.

[31] See Lee p. 84. The author states that the index of refraction of water in a cell is 1.3, with the molecules in solution in vacuoles is 1.34 and that cellulose of the wall has an index of refraction of 1.4.

with two components, parallel and perpendicular, then we can calculate the two components of the reflected component R and the refracted or transmitted component T. This is shown below:

$$T_0 = \frac{2}{n+1} A_0$$

$$T_1 = \frac{2}{n+1} A_1$$

$$R_0 = \frac{n-1}{n+1} A_0$$

$$R_1 = -\frac{n-1}{n+1} A_1$$

$$0 = parallel$$

$$1 = perpendicular$$

Thus when we enter the index of refraction we see that transmission is greater than reflection for perpendicularly incident light. This analysis provides some insight into the nature of the cell. Also one must remember that the reflected light is reflected off the outer layer of the cell. This process then continues again as the light leaves the cell and enters ones below.

We now move to a deeper analysis of the specific colorants and their impacts on the light entering the cells.

5.6 Colorants from Secondary Pathways

The anthocyanin molecules is shown below. Note on the B ring we have six sites to which we can attach differing molecular chains. This will be an important element when we see the different configurations and their implications.

5.6.1 Anthocyanins

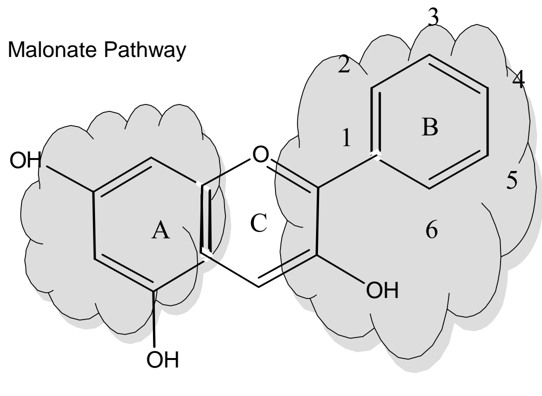
The anthocyanin or anthocyanidin molecule comes from two different secondary pathways in the plant cell. One is from the shikimic pathway and the other from the malonate pathway. This means that we have to understand both pathways to understand the ultimate abundance of the product. Anthocyanins are not the only elements which are secondary products which produce color. There are three classes of chemicals which give rise to color; anthocyanins, flavones or flavonols, and carotenoids. The basic structure of the anthocyanin is shown below.

Anthocyanidin

3
2
4
1
B
OH
O
A
C
5
6
OH
OH

The anthocyanin or Anthocyanidin molecules comes from two different pathways. In the figure below we have taken the basic resulting molecule and have shown that there are two elements; one is from the shikimic pathway and the other from the malonate pathway. This means that we have to understand both pathways to understand the ultimate abundance of the product.

Anthocyanidin



Before continuing we want to look at what the results would look like if we have different substitutes on the B ring. In the Table below we show that the terminations on the 3, 4 or 5 elements yield different results. The results give pelargonadin, cyanidin, delphinidin, peonidin, and petunidin. Each obviously named after their related flower and each resulting in an anthocyanin of a different color.

Anthocyanin Colors

Anthocyanidin	*Substituents*	*Color*
Pelargonidin	4'-OH	orange-red
Cyanidin	3'-OH, 4'-OH	purplish red
Delphinidin	3'-OH, 4'-OH, 5'-OH	bluish purple
Peonidin	3'-OCH_3, 4'-OH	rosy red
Petunidin	3'-OCH_3, 4'-OH, 5'-OCH_3	purple

6/5/2008 40

In the Table below we have shown the colors of each of these as well as the weighting of a red, green and blue combination which best matches the color. Thus one can in an 8 bit color schemes, as one would find in any PC color scheme, get the resulting anthocyanin colors by blending the R, B, G elements to yield what we are seeking. This relating the colors back to RBG is critical since it get reflected in the ultimate flower color.

Colors (R, G, B)

Pelargonadin (255, 102, 0)	
Cyanidin (255, 0, 255)	
Delphinidin (153, 102, 255)	
Peonidin (255, 153, 204)	
Petunidin (153, 0, 153)	

Now if we assume we have only anthocyanins for color, and that we have the above combinations available, we ask; how do we combine these colors in a weighted manner to obtain the desired color? This approach is critical to the overall understanding. First we show by a weighted RBG we get the color we seek or the color which is presented. Then we assume that if we can then do the same for each anthocyanin, then we can create any desired color from a weighted collection of anthocyanins. This means that we can then determine what the relative percents of expression of any anthocyanin is and

this lets us then go back to how strongly the gene for that anthocyanin is expressed. The model we presented earlier will be a key element in this overall process.

No let us start with a simple expression. For any color we have by definition:

$$\text{Color} = \alpha\,\text{Red} + \beta\,\text{Blue} + \delta\,\text{Green}$$

For example, we may have a (0,0,255), or a (128, 128, 128). Or any other set of combinations.

Likewise we could state this by means of some combination of anthocyanins. Namely:

$$\text{Color} = \text{a}\,[\text{Pelargonidin}] + \text{b}\,[\text{Cyanidin}] + \text{c}\,[\text{Delphinidin}]$$

But we can relate the anthocyanins to the basic colors or red, blue and green as:

$$[Pel\arg onidin] = \alpha_P \operatorname{Re} d + \beta_P Blue + \delta_P Green$$

$$[Cyanidin] = \alpha_C \operatorname{Re} d + \beta_C Blue + \delta_C Green$$

$$[Delphinidin] = \alpha_D Red + \beta_D Blue + \delta_D Green$$

If we define a color vector of Red, Blue and Green as:

$$C = \begin{bmatrix} Red \\ Blue \\ Green \end{bmatrix}$$

Thus if we define the mix vector as **m** then we have:

Color = $\underline{\mathbf{m}}^T \underline{\mathbf{C}}$

Or:

$$m = \begin{bmatrix} \alpha \\ \beta \\ \delta \end{bmatrix}$$

But we have the following matrix:

$$A = \begin{bmatrix} \alpha_P & \beta_P & \delta_P \\ \alpha_C & \beta_C & \delta_C \\ \alpha_D & \beta_D & \delta_D \end{bmatrix}$$

These yields:

$$\text{Color} = \underline{\mathbf{m}}^T \underline{\mathbf{A}}\, \underline{\mathbf{C}}$$

The above analysis shows us that we can analytically determine the expression of the anthocyanins from the color of the cell by means of the above formulas. These are relative expressions but by benchmarking any one element we can make them all absolute in the cell as well.

5.6.2 OTHER COLOR ELEMENTS

Anthocyanins are not the only elements which are secondary products which produce color. There are three classes of chemicals which give rise to color; anthocyanins, flavones or flavonols, and carotenoids. The Table below depicts the different elements and their colors. The approach we took above for the anthocyanins can be taken for the flavones and carotenoids as well. It should be noted that there may not be a unique solution here but there are several possible but they can be narrowed down by actual determination of one to three elements as baseline.

Class	*Agent*	*Color*[32]
Anthocyanidin		
	Pelargonidin	orange-red
	Cyanidin	purplish-red
	Delphinidin	bluish-purple
	Peonidin	rosy red
	Petunidin	purple
	Malvinidin	
Flavonol		
	Kaempferol	ivory cream
	Quercetin	cream
	Myricetin	cream
	Isorhamnetin	
	Larycitrin	
	Syringetin	
	Luteolin	yellowish
	Agipenin	Cream
Carotenoids		
	Carotene	orange
	Lycopene	Orange-red

We now summarize the other element classes.

5.6.2.1 Carotenoids

Carotenoids are what is quite common in the carrot, the orange hew we see in that root. Its molecular structure is shown below, this is beta carotene.

[32] See Taiz p. 334 for the Anthocyanidin color and Bernhardt for the Flavonol and carotene.

5.6.2.2 Flavones

The flavonols, or flavones are quite similar to anthocyanin. Their structure is shown below. Note that we have compared it to that of anthocyanin.

Anthocyanidin

Flavonol

We can also show how closely they relate in substitutions and colors. This is shown in the Table below.

Flavonol	Anthocyanidin	Substitution	
		3’	5’
Kaempferol	Pelargonidin	H	H
Quercetin	Cyanidin	OH	H
Myricetin	Delphinidin	OH	OH
Isorhamnetin	Peonidin	OCH_3	H
Larycitrin	Petunidin	OCH_3	OH
Syringetin	Malvinidin	OCH_3	OCH_3

5.6.3 ANTHOCYANIN ABSORBANCE AND REFLECTANCE

There have been many studies on the absorbance and reflectance of the various anthocyanins. The Figure below shows the results from a 1957 paper by Harborne. Harborne used the following procedure:

"*Spectrophotometry. All measurements were made with a Unicam SP. 500 spectrophotometer. The pure dry pigments were dissolved in methanol containing 0-01 % of conc. HCI and the solutions diluted to give an optical density reading in the range 0-800-1-300 at the visible maxima. For measurements in the ultraviolet region, the solutions of those anthocyanins obtained from eluting chromatograms were measured against a solution obtained from an appropriate blank area of the chromatogram, prepared at the same time as the corresponding pigment solution.*

For the purpose of measuring spectral shifts in the presence of aluminum chloride, three drops of a solution of the anhydrous salt in ethanol (5%, w/v) were added to the cell solution. Measurements of the shift were made as quickly as possible, since in some cases the color of the resulting solution faded on standing.

For measuring the spectra of mixtures of Pelargonidin 3:5-diglucoside and p-coumaric, caffeic and ferulic acids, 1 m-mole of each compound in pure, anhydrous form was dissolved in 50 ml. of methanol containing 0-01% of conc. HCI. Portions (1 ml.) of these stock solutions were mixed together in varying proportions and the solution was made up to 10 ml. The compounds were previously purified by recrystallization and then dried in vacuo at 1000 over phosphorus pentoxide."

He continues:

"In searching for new methods of characterizing anthocyanins, it should be remembered that they are difficult compounds to deal with by the usual techniques of organic chemistry. In solution, they are unstable to light and pH changes. They are difficult to isolate in a pure state as a general rule. Many of them do not have sharp melting points and do not give meaningful results on elementary analysis. No suitable derivatives are known for characterizing them. The procedure of methylation and hydrolysis, commonly used with flavones for determining the position of sugar residues, is of limited value in the anthocyanin series..."

The Harborne result for Pelargonidin is shown below.

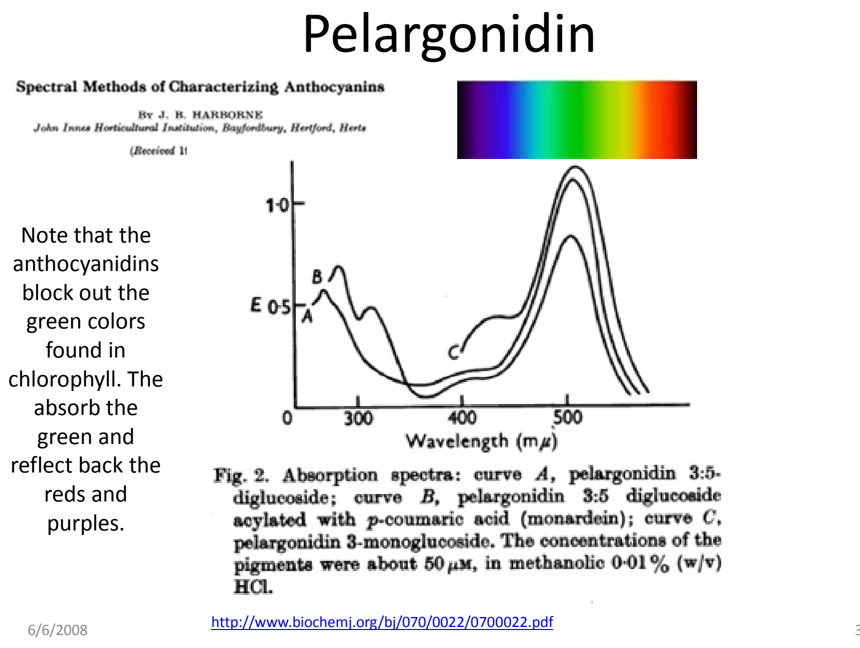


Fig. 2. Absorption spectra: curve *A*, pelargonidin 3:5-diglucoside; curve *B*, pelargonidin 3:5 diglucoside acylated with *p*-coumaric acid (monardein); curve *C*, pelargonidin 3-monoglucoside. The concentrations of the pigments were about 50 μM, in methanolic 0·01 % (w/v) HCl.

The following Table details all of the anthocyanins and their peak spectral line as well as the relative peaks compared to 440 nm[33].

[33] From Harborne, 1958.

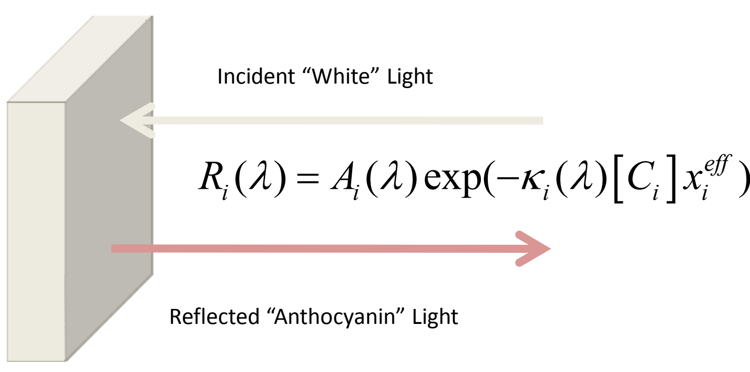
Anthocyanidin Class	Anthocyanidin Specific	Wavelength nm	E440/Emax
Pelargonidin and derivatives	Pelargonidin Pelargonidin 3-monoglucoside Pelargonidin 3-rhamnoglucoside Pelargonidin 3-gentiobioside Pelargonidin 3-diglucosido-7 (or 4'.)-glucoside	520 506 508 506 498	39 38 40 36 42
Cyanidin and derivatives	Cyanidin Cyanidin 3-monoglucoside Cyanidin 3-rhamnoglucoside Cyanidin 3-gentiobioside Cyanidin 3-xyloglucoside Peonidin Peonidin 3-monoglucoside	535 525 523 523 523 532 523	19 22 23 25 22 25 26
Delphinidin and derivatives	Delphinidin Delphinidin 3-monoglucoside Delphinidin 3-rhamnoglucoside Petunidin Petunidin 3-monoglucoside Malvidin With 5-hydroxyl group free Malvidin 3-monoglucoside With 5-O-substituent	544 535 537 543 535 542 535	16 18 17 17 18 19 18
Pelargonidin and derivatives	Pelargonidin 5-glucoside Pelargonidin 3:5-diglucoside Pelargonidin 3-rhamnoglucosido-5.glucoside Pelargonidin 3-diglucosido-5-glucoside Monardein Salvianin	513 504 505 503 505 505	15 21 19 21 21 20
Cyanidin and derivatives	Cyanidin 3:5-diglucoside Peonidin 3:5-diglucoside Peonidin 3-rhamnoglucosido-5-glucoside Peonidin 5-glucoside Peonidin 5-benzoate	522 523 523 528 528	13 13 12 12 11
Delphinidin and derivatives	Delphinidin 3:5-diglucoside Petunidin 3:5-diglucoside Petunidin 3-rhamnoglucosido-5-glucoside Malvidin 3:5-diglucoside Malvidin 3-rhamnoglucosido-5-glucoside Negretein	534 533 535 533 534 536	11 10 10 12 9 9

5.7 Estimating Anthocyanin Concentrations

This section addresses the ability to determine the detailed concentrations of each of the colorants in a cell if one knows the cell effective optical length and the extinction coefficients for each of the constituents. The models for performing these tasks also show what the maximum resolution that can be achieved as well and the maximum number of constituents. The results in the maximum bounding resemble the same results that are found in such areas as ascertaining the accuracy in ambiguity functions for phased arrays. The latter problem was solved by the author in the mid 1970s.

5.7.1 The Model

Let us begin with a simple model of reflectance. We look at the Figure below and see a white light impinging on a cell and the light reflected back is seeing at one specific wavelength, frequency, as an attenuated version of what was transmitted at the wavelength. A is the amplitude of the transmission and the exponentially reduced A value is what is reflected. Thus if absorption is in the red and blue as we saw with chlorophyll then we reflect green and that is what we see. This is an application of Beer's Law[34]. Beer's law is a statistical approach to absorption. It reflects what experimentally is obtained and does not provide a detailed analysis as we had been developing in prior sections.



We define the reflect light at a specific frequency, wavelength, as follows:

$$R_i(\lambda) = A_i(\lambda)\exp(-\kappa_i(\lambda)[C_i]x_i^{eff})$$

where [C] is a concentration and x is the effective thickness of the cell.

Here R is the reflected light we see at the wavelength specified and at the ith anthocyanin.

---

[34] See Cantor and Schimmel, pp. 60-68.

A is the incident light amplitude at the wavelength specified. The exponent is Beer's law where C is the concentration of anthocyanin I and x the effective depth of that anthocyanin.

Now we can write Beer's law for one or two or even more absorbents. We show the case for one and two absorbents as follows:

$$\frac{dR}{R} = -C_k \kappa_k dx$$
$$or$$
$$\frac{dR}{R} = -C_k \kappa_k dx - C_{k+1} \kappa_{k+1} dx$$

Note that the reduction in reflected light or in transmitted light is reduced by a result of the additive reduction of separate collisions with separate molecules.

$$R_{Total}(\lambda) = A(\lambda)\exp(-\sum_{i=1}^{N} \kappa_i(\lambda)[C_i]x_i(\lambda))$$

The log of the ratio of intensities is the sum of the weighted concentrations. We assume we know the κ values for each absorbing element at each wavelength. Then we can use the above to estimate the separate concentrations

$$I(\lambda) = \ln \frac{R_{Total}(\lambda)}{A(\lambda)} = -\sum_{i=1}^{N} \kappa_i(\lambda)[C_i]x_i(\lambda)$$

The problem is simply stated. We measure the intensity at say M values of wavelength and this gives us M samples. We then must find values of the [C] which give the best fit to the measurements obtained using the model assumed. That is for every wavelength, we define an error as the difference between the measurement and what the measurement would have been using the estimates of the [C] values and the best [C] values are those which minimize the sum of the squares of these errors. There are M measurements and N concentrations and M is much larger than N. That is:

$$\text{Choose } [C_n] \text{ such that they minimize}$$

$$\min(\sum_{m=1}^{M}\left(I(m)-\hat{I}(m)\right)^2$$

where

$I(m)$ *is the mth measurement*

and

$$\hat{I}(m)=\sum_{i=1}^{N}\kappa_i(\lambda)[\hat{C}_i]x_i(\lambda)$$

This is an optimization problem which can be solved in many ways. We address some of them in the next section.

5.7.2 The Approaches

Some Examples, this is an example of the Inverse Problem already solved by McGarty:

1. CIE approach: This assumes that one can unravel the exponents of the x,y,z model. The problem is that we will not have an adequate number of degrees of freedom.
2. Splines: This assumes we can generate curves and then separate them and then focus on their coefficients[35].
3. Steepest Descent: This is the incremental approach of best fit. It assumes we are trying to solve an optimization problem.
4. Least Squares Fit: A statistical best fit method.
5. Kalman Filter: This is the statistical solution using steepest descent but with correlation matrices.
6. Matched Filter: This approach assumes we know the waveforms of each absorption curve for each colorant and that we receive a resulting absorption curve which is the sum of all of them, and that we then try to estimate the "amplitudes" of each curve, in effect the concentrations.

We may define the problem as follows:

[35] See Hildebrand pp. 478-494. The use of splines is an approach which tries to match coefficients of polynomials.

Let $R(\lambda)$ be determinable for a given set of $[C_i]$ and let
$\widetilde{R}(\lambda)$ be the measured received spectrum power and
$I(\lambda)$ be the log of the received to incident power
at the wavelength
Find the set of $[\widehat{C}_i]$, i=1...N, such that
$\left(\widetilde{R}(\lambda)-\widehat{R}(\lambda)\right)^2$ is minimized where $\widehat{R}(\lambda)$is the
estimated received spectral element

We may also characterize the variables as follows:

Let

$$x(k)=\begin{bmatrix} C_1 \\ . \\ . \\ C_n \end{bmatrix}=x(k+1)$$

and

$$z(k)=c^T(k)x(k)+n(k)$$

where

$$c(k)=\begin{bmatrix} -\kappa_1(k)x_1 \\ . \\ . \\ -\kappa_n(k)x_n \end{bmatrix}$$

and for this case k and λ are identical increments

We now consider three possible approaches.

5.7.2.1 Newton Steepest Descent

The Newton Steepest descent approach is one where we define an optimization and this optimization results in solving a polynomial equation. We then employ an iterative method to solve that equation. We now seek the following:

Find the a such that:

$$\hat{a} = \begin{bmatrix} \hat{a}_1 \\ .. \\ \hat{a}_n \end{bmatrix} = \begin{bmatrix} [C_1] \\ . \\ . \\ [C_N] \end{bmatrix}$$

such that

$$\min\left[\sum_{i=1}^{M} (I_i - \hat{I}_i)^2 \right]$$

Let us recall the simple optimization result:

$$h(a) = \left[\sum_{i=1}^{M} (I_i - \hat{I}_i)^2 \right]$$

and

$$\frac{\partial h(a)}{\partial a_n} = g_n(a) = 0$$

is the optimal point, so we seek to solve the vector equation:

g(a)=0

We can now state the general solution in terms of Newton's Method[36]:

$$g(a) = 0$$

is the desired result. Define:

$$A(a) = -\left[\frac{\partial g(a)}{\partial a} \right]^{-1}$$

where we define:

$$\left[\frac{\partial g(a)}{\partial a} \right] = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} \cdots\cdots \frac{\partial g_1}{\partial a_n} \\ \frac{\partial g_n}{\partial a_1} \cdots\cdots \frac{\partial g_n}{\partial a_n} \end{bmatrix}$$

and the estimate at sample k+1 is:

$$\hat{a}(k+1) = \hat{a}(k) + A(\hat{a}(k))g(\hat{a}(k))$$

[36] See Athans et al, Systems, Networks and Computation, Multivariable Methods, McGraw Hill (New York) 1974, pp-115-122.

Note that we use this iterative scheme as one of several means to achieve the result. For each tuple of data we do the following:

$\hat{a}(0) = a^0$, an n x 1 vector guess. Then we use the first data tuple:

$\hat{a}(1) = \hat{a}(0) + A(\hat{a}(0))g((\hat{a}(0))$

where we use the difference:

$a_{k,measured}(0) - \hat{a}(0)$

as the data entry element for each of the elements of a.

The Newton algorithm is but one of many possible algorithms. We know the conditions for Newton convergence. We can also estimate the accuracy of this algorithm as well.

5.7.2.2 KALMAN FILTER

The method of estimating the structural elements of the gene expression can be structured using a standard set of methodologies. In particular we use the two approaches. The approach was applied to estimating the constituent chemical concentrations of the upper atmosphere, namely the inversion problem, using transmitted light as the probe mechanism. In this case we seek to estimate the gene expression matrix using the concentrations of secondary chemicals as expressed in color concentrations. This is in many ways a similar problem.

Let us consider a six gene model, two color modifying genes and four control genes, two each. The model is as follows. First is a general linear model for the gene production:

$$\frac{dx(t)}{dt} = Ax(t) + u(t) + n(t)$$

Then the entries are as follows:

$$A = \begin{bmatrix} a_{11}..a_{12}..a_{13}..0..0..0 \\ 0...a_{22}..0...0...0...0 \\ 0...0...a_{33}..0...0...0 \\ 0...0...0...a_{44}..a_{45}..a_{46} \\ 0...0...0...0...a_{55}..0 \\ 0...0...0...0...0...a_{66} \end{bmatrix}$$

and

$$u(t) = \begin{bmatrix} u_1 \\ ... \\ u_6 \end{bmatrix}$$

And we assume a system noise which is white with the following characteristic:

$$E[n(t)] = 0$$
and
$$E[n(t)n(s)] = N_0 I \delta(t-s)$$

Now we can define:

$$A = \begin{bmatrix} A_1 ... 0 \\ 0 ... A_2 \end{bmatrix}$$

Where we have partitioned the matrix into four submatrices. This shows that each gene and its controller are separate. Now we can determine the concentrations of each protein in steady state as follows, neglecting the Gaussian noise element for the time being:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = -A_1^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$
and
$$\begin{bmatrix} x_4 \\ x_5 \\ x_6 \end{bmatrix} = -A_2^{-1} \begin{bmatrix} u_4 \\ u_5 \\ u_6 \end{bmatrix}$$

We argue that finding either the matrix elements or their inverse relatives is identical. Thus we focus on the inverse elements. Now the concentrations of the anthocyanins are given by the 2 x 2 vector as follows:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} c_{11} ... 0 ... 0 ... 0 ... 0 ... 0 \\ 0 ... 0 ... 0 ... c_{24} ... 0 ... 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = Cx$$

The color model remains the same.

5.7.2.2.1 The Estimator Model

The system model is as follows. Let us begin with a model for the vector a that we seek:

$$\frac{da(t)}{dt} = 0 : where$$

$$a(t) = \begin{bmatrix} a_1 \\ \dots \\ a_5 \end{bmatrix}$$

In this case we have assumed a is a 5 x 1 vector but it can be any vector. The measurement system equation is given by:

$$z(t) = g(a,t) + w(t)$$

Where z is an m x 1 vector. In this case however we have for the measurement the following:

$$z(t) = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ x_1 \\ \dots \\ x_6 \end{bmatrix} = g(a,t) + w(t)$$

We now expand in a Taylor series the above g function:

$$g(a,t) = g(a_0,t) + C(a_0,t)\left[a(t) - a_0(t)\right] +$$
$$\frac{1}{2}\sum_{i=1}^{N} \gamma_i \left[a - a_0\right]^T F_i \left[a - a_0\right] + \dots$$

Where we have:

$$C = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \dots & \frac{\partial g_1}{\partial a_n} \\ .. & .. & .. \\ \frac{\partial g_m}{\partial a_1} & \dots & \frac{\partial g_m}{\partial a_n} \end{bmatrix}$$

Thus we have for the measurement:

$$z(t) = C(t)a(t) + \left[g(a_0) - C(a_0)a_0(t)\right]$$

We now use standard Kalman theory to determine the mean square estimate;

$$\frac{d\hat{a}(t)}{dt} = P(t)C^T(t)K^{-1}(z - C(t)\hat{a}(t))$$

where

$$\frac{dP(t)}{dt} = -P(t)C^T(t)K^{-1}C(t)P(t) +$$

$$\sum_{i=1}^{N} PF_i P\gamma_i{}^T K^{-1}(z - g(a_0)$$

where

$$K\gamma(t-s) = E\left[w(t)w^T(s)\right]$$

In discrete time we have the equation:

$$\hat{a}(k+1) = \hat{a}(k) + PCK^{-1}\left[z(k) - \hat{z}(k)\right]$$

This is identical to the equation we derived from the Newton method.

5.7.2.3 The Matched Filter Approach

This is a different approach and it is an application of signal detection taken from classic communication theory. It assumes we have N signals and each signal shape is known but the amplitude of the individual signals is not known. Then we ask how we can estimate the amplitude of each signal if what we have is a received signal which is the sum of the N plus noise. We begin this approach as follows:

Let us assume there are two waveforms bounded on an interval [0, T]

Let

$$s_1(t) = s_1^{orthog}(t) + s_1^{remain}(t)$$

$$s_2(t) = s_2^{orthog}(t) + s_2^{remain}(t)$$

such that

$$\int_0^T s_1^{orthog}(t) s_2^{orthog}(t) dt = 0$$

Now there are three questions which we may pose:

1. Does such a decomposition exist, if so under what terms?

2. What is a constructive way to perform the decomposition?

3. Is there an optimum decomposition such that the "distance between the two orthogonal signals is maximized"?

Namely:

$$\exists\ a\ set\ \{s_1, s_2\}$$
$$such\ that$$
$$\int_0^T s^2{}_1(t)dt = E_1$$
$$\int_0^T s^2{}_2(t)dt = E_2$$
$$and$$
$$\exists \max$$
$$\int_0^T s^{2,orth}{}_1(t)dt = \widetilde{E}_1$$
$$\int_0^T s^{2,orth}{}_2(t)dt = \widetilde{E}_2$$

Let us approach the solution using the theory of orthogonal functions[37]. Now we can specifically use a Fourier series approach. We do the following:

$$Let$$
$$s_1(t) = \sum_{n=1}^{\infty} s_1^n \cos(\frac{2\pi}{T} nt) + r_1(t)$$
$$where$$
$$r_1(t) = s_1(t) - FS\cos$$
$$and$$
$$FS\cos = \sum_{n=1}^{\infty} s_1^n \cos(\frac{2\pi}{T} nt)$$

Likewise

[37] See Sansone, Orthogonal Functions.

$$
\begin{aligned}
&Let \\
&s_2(t) = \sum_{n=1}^{\infty} s_2^n \sin(\frac{2\pi}{T} nt) + r_2(t) \\
&where \\
&r_2(t) = s_2(t) - FS\sin \\
&and \\
&FS\sin = \sum_{n=1}^{\infty} s_2^n \sin(\frac{2\pi}{T} nt)
\end{aligned}
$$

Clearly FScos and FSsin are orthogonal. The residual functions r are the sin and cos elements respectively of the expansions. We could have just as easily transposed the sin and cos allocations between the two s functions. As to answering the third question we are effectively asking if the r residual functions can be minimized. The answer is not with a Fourier Transform. Then the question would be; is there another set of orthogonal functions which would minimize the residuals, namely:

$$
\begin{aligned}
&\int_0^T r_1^2(t)dt = R_1 \\
&and \\
&\int_0^T r_2^2(t)dt = R_2
\end{aligned}
$$

are to be minimized. For a Fourier Transform as the orthogonal base we are left with residuals, R, at whatever they may be. However using the Fourier Transform approach we can extract the two signals as follows:

$$
P(\lambda) = [C_1]\tilde{s}_1(\lambda) + [C_2]\tilde{s}_2(\lambda) + r_{TOT}(\lambda)
$$

Note we can interchange t and λ since they represent the same variable. We now have a "signal" with amplitudes to be determined and a bias which is known. Using standard "signal detection theory" we can readily solve this problem as well. This becomes the "matched filter problem"[38].

5.8 CONCLUSIONS

What we have sought to accomplish in this Chapter is to describe color and it generation in plants and to present a set of methods and means to determine the constituents which give rise to those colors. In effect we have created a world view of color, apart

[38] See VanTrees, Detection, Estimation and Modulation Theory. He presents details on this solution.

from the classic colorimetry approach, and used this and the physical measurements related thereto to affect a method and means to determine concentrations of colorants in flowers.

The simple application of Beer's law and the use of the known spectra of anthocyanins and other colorants allow us to use data from FTS to determine the concentrations of each colorant on literally a cell by cell basis. Beer's law is a simplistic but fairly accurate and consistent method. It would be interesting to explore the details of the transmission of light to a deeper level but the complexity of the cell structure prohibits that at this time.

Having a methodology of the type developed herein we can now more readily examine the genetic pathways and expression systems in the genus Hemerocallis. This paper details multiple ways to ascertain concentrations on a cell by cell basis.

6 FLOWER COLOR, PATTERNING AND CONTROL

This Chapter deals with two issues. First we address the issue of patterning in flowers. Namely we use the Turing model for tessellation to show how the patters we see on the recent daylilies are formed. This analysis is then looked at in terms of actual patterned flowers and it is demonstrated that the model yields reliable results. The second issue we address is the actual genetic processes and control of secondary pathways which control color in a cell. We develop a detailed model which permits the identification of the control processes and secondly permits the development of processes that can generate any desired color. The first discussion is the inter-cell discussion and the second issue focuses on the inter-cell issue. The combination of these two analyses effectively established a genetically controlled framework for color and patterning in flowers. In fact this is extensible to any genetically controlled and controllable secondary pathway processes. To the author's knowledge, there has been no similar analysis of the form presented in this Chapter.

6.1 PATTERNING

The development of patterns in flowers has been examined by many over the years. The seminal work of Alan Turing in 1952 laid out a method to model such coloration by looking at the process as a distributed communications between cells with feedback. This paper uses the Turing model combined with the current knowledge of gene expression and secondary pathways. We look at the patterns as classified by Petit as a canonical baseline and from the Petit Patterns we apply the Turing-Murray model to validate that the patterns can be duplicated by well defined and reproducible genetic control mechanisms.

The production of patterns in flowers has been an intriguing process with limited explanation. In this paper we propose a model for analyzing that process, for experimentally verifying the process, and for being able to reverse engineer the process via control of genetic pathways. There has been a great deal of discussion in the Hemerocallis community regarding the genetics of the flower. There are two issues: (i) what are the true species and how are they defined, (ii) what causes the coloration and does one have a genetic explanation for them. In this paper we continue our work on the second question. There has been earlier works by Norton who proposed asset of genes, without any evidence, that control colors. This was a classic Mendellian approach. The Norton model fails to deal with the known pathways generating the colorants such as anthocyanin and totally fails to relate that to the now know gene enzyme pathway controlling products. This paper provides an integral approach which is experimentally verifiable to explain and obtain patterned flowers. As such our approach herein uses in each step experimentally verified or verifiable procedures to explain patterning.

6.1.1 Prior Work on Patterning

We begin with a summary of the work performed to date.

1. Genes and Gene Control: The understanding of the gene and its functions began with the publication of the Watson and Crick paper in April 1953. In August 1952 the Royal Society published a paper by Alan Turing entitled the Chemical Basis for Morphogenesis. In this paper Turing proposed a solution to the color pattern problem but unfortunately he had to hypothesize chemical reactions which no one at the time was yet aware. In the 1953 paper by Watson and Crick, the authors proposed a structure for DNA and they also proposed the mechanism for DNA making RNA and then proteins. This was the beginning. After some 55 years we now know a great deal about the mechanics of this system.

2. Plant Color and Anthocyanins: Plant colors, especially the flowers, are controlled by such secondary chemicals as anthocyanins. The anthocyanins absorb from the white incident light and reflect the colors we perceive in the flower. The secondary colorants such as the anthocyanins result from secondary chemical pathways which are driven by enzymes, proteins, produced by genes in the cell. The recent work by Mol et al, Jaakola, and Winkle-Shirley has provided reviews which provide up to data understanding of these processes. Each cell has its own secondary colorant process and the resulting concentrations of the colorants create a cell by cell color. Each cell may therefore result in differing concentrations of colorants and these concentrations are controlled by the genetic pathways in the cell, and are also affected by the flow of genetic pathway proteins which may have arrived from adjacent cells as well. The recent work by Mol (1998, 1999) Winkel Shirley (2001) and the work of Milgrom provide details on these processes. Likewise that of Durbin provides a connection with the evolution of the plants based on color structure. Harborne (1958, 2001) has provided a detailed basis to analyze the anthocyanins by spectral methods and we shall use these in this analysis. The other pathways such as the carotenoids have been recently summarized by Naik. A summary of all pathways has also recently been done by Yu and also by Holton et al. as well as the recent book by Lee. We rely heavily on the Mol and Winkel Shirley work.

3. Enzyme Reactions and Control: Proteins generate by the plant genes are the enzymes which effect the production and production rates of the secondary pathway colorants. Recent review papers by Baici shows the variants of models which can be applied to the dynamics here. Enzymes take part in the reactions by acting as facilitators. The more enzymes present the more the reaction moves forward. The enzymes may be produced locally in the specific cell or they may flow into the cell from adjacent cells. In addition the work by Murrells also provides a strong basis for modeling the reaction kinetics in this approach.

4. Intercellular Flow of Proteins: Plant cells have intercellular communications paths which differ greatly from those of animals. The paths are facilitated by the plasmodesmata. Recent studies have greatly elucidated the operations of these elements of plants and have shown that they are an integral part in the control of the overall genetic pathways and secondary pathways in plants. The recent work summarizing this field by Haywood et al, Cilia et al and Oparka and Zambryski display the extensive knowledge of the plasmodesmata and its role in the control in gene expression across a large matrix of plant cells. As we shall see latter, this mechanism for intercellular transfer can be viewed as the basis for a diffusion process between cells.

5. Patterning in Plants: In 1952 Alan Turing, in the last year and a half of his life, was focusing on biological models and moving away from his seminal efforts in encryption and computers. It was Turing who in the Second World War managed to break many of the German codes on Ultra and who also created the paradigm for computers which we use today. In his last efforts before his untimely suicide Turing looked at the problem of patterning in plants and animals. This was done at the same time Watson and Crick were working on the gene and DNA. Turing had no detailed model to work with, he had no gene, and he had just a gestalt, if you will, to model this issue. Today we have the details of the model to fill in the gaps in the Turing model.

6.1.2 THE TURING MODEL

The Turing model was quite simple. It stated that there was some chemical, and a concentration of that chemical, call it C, which was the determinant of a color. Consider the case of a zebra and its hair. If C were above a certain level the hair was black and if below that level the hair was white. As Turing states in the abstract of the paper:

"It is suggested that a system of chemical substances, called morphogens, reacting together and diffusing through a tissue, is adequate to account for the main phenomena of morphogenesis. Such a system, although it may originally be quite homogeneous, may later develop a pattern or structure due to an instability of the homogeneous equilibrium, which is triggered off by random disturbances. Such reaction-diffusion systems are considered in some detail in the case of an isolated ring of cells, a mathematically convenient, though biologically unusual system. The investigation is chiefly concerned with the onset of instability. It is found that there are six essentially different forms which this may take. In the most interesting form stationary waves appear on the ring. It is suggested that this might account, for instance, for the tentacle patterns on Hydra and for whorled leaves. A system of reactions and diffusion on a sphere is also considered. Such a system appears to account for gastrulation. Another reaction system in two dimensions gives rise to patterns reminiscent of dappling. It is also suggested that stationary waves in two dimensions could account for the phenomena of phyllotaxis.

The purpose of this paper is to discuss a possible mechanism by which the genes of a zygote may determine the anatomical structure of the resulting organism. The theory does not make any new hypotheses; it merely suggests that certain well-known physical laws are sufficient to account for many of the facts. The full understanding of the paper requires a good knowledge of mathematics, some biology, and some elementary chemistry. Since readers cannot be expected to be experts in all of these subjects, a number of elementary facts are explained, which can be found in text-books, but whose omission would make the paper difficult reading."

Now, Turing reasoned that this chemical, what he called the morphogen, could be generated and could flow out to other cells and in from other cells. Thus focusing on one cell he could create a model across space and time to lay out the concentration of this chemical. He simply postulated that the rate of change of this chemical in time was equal to two factors; first the use of the chemical in the cell, such as a catalyst in a reaction or even part of the reaction, and second, the flow in or out of the cell. The following equation is a statement of Turing's observation.

$$\frac{\partial C_1(x,t)}{\partial t} = F_1(C_1, C_2, x, t) + \lambda_1 \nabla^2 C_1(x,t)$$

This is a nonlinear diffusion equation. It allows one to determine the concentration of say a protein as a function of time and space across a cell matrix. It requires two fundamental experimental parameters: first, the factors in the rate portion of the equation frequently found in reaction kinetics and second a diffusion constant determined by intercell transport most likely via the plasmodesmata. Recent works by Benson and Benson et al have added additional structure to the approach as well as the text by Murray establishing a full framework for the analysis of the model.

In the model we develop herein, we look at the concentrations of activator and repressor protein concentrations generated by activator and repressor genes. We employ the Murray-Turing model which states that the gene controlling a specific anthocyanin pathway is either on or off depending upon the concentration of activator or repressor. If activator concentration exceeds the repressor concentration then the gene functions and drives the secondary pathways thus producing the desired anthocyanin. In a linearized model of this process all one needs is linearized reaction coefficients and diffusion constants. The result is a two dimensional wave of activator and repressor concentrations and the result of these two waves is a set of fluctuating color patterns. If we were to then do this for multiple anthocyanin pathways we would then determine a set of multiple patterns which have been described by Petit.

6. Genetic Expression and Secondary Pathways, an Integrative View: The next step in developing this approach is to have an experimentally established model for the control of secondary pathways. We know from prior work that the secondary pathways

exist and that they produce anthocyanins. The anthocyanins are the basis for plant color. We have established elsewhere that one can determine concentrations of anthocyanins observationally from a cell by cell spectroscopy (see McGarty, 2008). Namely there is an inversion process which permits the estimation of concentration densities from spectroscopic data.

The gene expression modeled developed in McGarty (2007) shows how the genes which create the protein which is the enzyme regulating each of the colorants pathways are themselves controlled by repressor and/or activator genes. By means of microarray analysis it is possible to both identify those genes and to determine the degrees of coupling between them as well.

Thus from the work of McGarty (2007) we have a model that connects genes to secondary pathways and moreover allows the connection to be quantified.

7. Patterning and a Canonical Model: Patterns in flowers have been a driving factor for many hybridizers who seek to have unique plants for sale. Ted Petit is a Hemerocallis breeder who is well known for his patterned flowers. He has recently published an article which has place many of the patterns in some reasonable analytical form (see Petit, 2007). The canonical forms proposed by Petit, we refer to them as "Petit Patterns", cover a wide gamut of the flower structures observed in Hemerocallis. One can then takes the Petit Patterns and using the Turing Space analysis discussed above and performs an experimental verification.

Turing did not have many of the elements we have developed above. He did not even know of the gene as we now know it, for he died less than a year after Watson and Crick published their famous paper. In fact, Turing went on trial for his admitted indiscretions merely days after he published his paper. At the time of his work many saw him as attempting to describe how daisies have so complex a petal set and the like. One may wonder what he would have thought of the problem as posed by Petit.

7. The selection of Hemerocallis as a target genus is based upon the fact that a great deal is known about the species but more importantly the hybridization of the genus over the past hundred years provides a great genetic pool to track the development of color and patterns. The original work on the genus was written by Stout in the 1930s and was reissued in the late 1980s. There is a great deal of field and analysis work performed in Japan, Korean and some in China as well. The work by Munson depicts the status of the genus as it has been hybridized until the early 1990s. Kang and Chung have performed many studies looking at the genus from a systematics perspective as well as a genetic perspective. Erhardt has proposed a key which he alleges is better than that of Stout. Unfortunately Erhardt provides no justification for his phylogenetic key whereas the Stout key was used solely for identification and had no systematics function.

Tompkins has extended the analysis of the genus by using AFLPs and his analysis appears to support some but clearly not all of the alleged Erhardt phylogenetic analysis.

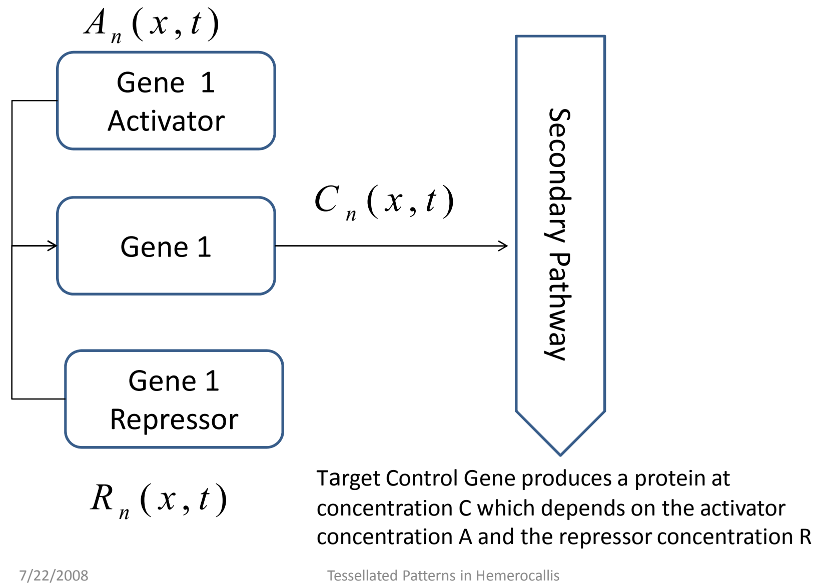
We first provide an overview and canonical model of the genetic pathways and secondary pathways which control color. The canonical model is shown below. The model is based upon the following experimentally observed facts:

1. There exist secondary pathways which convert primitive compounds into complex compounds which are the basis for color in plants. These compounds may be anthocyanins or many other similar secondary compounds.

2. There exists a gene which acts upon a secondary pathway as an enzyme in a catalytic manner. The concentration of the gene acting on the secondary pathway may increase the conversion in that pathway resulting in higher concentrations of the element activated. The greater the concentration of one secondary product as compared to another the more they viewed color may change. Secondary elements act as additive colorants in the Newtonian sense. They do not act as pigments which are subtractive. Care must be taken in calling the secondary elements pigments since their behavior in a cumulative manner is additive rather than subtractive as one finds in classic pigments in the world of painting.

3. There exist Activator and Repressor genes which can modulate the production of the target gene used to control, the secondary pathway. These Activator and Repressor genes may or may not themselves be so controlled. For the purpose of simplicity we assume at most one dominant Activator and Repressor gene.

We graphically depict this model below.



The processes in plant genes are generally identical to those in animal and thus human genes. The figure below shows a typical gene structure along with key sites. This structure shows the gene activator site which is where activator proteins can bind to start or enhance the expression of the gene.

6.1.3 Experimental Methods for Validation

We have obtained samples from many hybrids of the genus Hemerocallis and have sectioned the petals. The sectioned petals were obtained using standard microtome techniques and we focused on taking sections which were indicative of the bands of color variation.

We then mounted the sections on slides and used no other means of preparation.

We then used a Pentax 200D to capture all images in the RAW mode. We performed this for both the slide data as well as the inflorescence data using RAW format and doing so within a six hour period. The choice of the six hours was to ensure that the anthocyanins did not degrade. All petal and sepal samples were kept refrigerated and in the dark during the waiting period.

The capture resolution used was 8.6 Mb of resolution. Comparisons were also made with jpg capture modes and no significant difference was obtained.

We also took sections of slides of the flower and of the cells and using Photoshop CS3 determined the color content on a cell by cell basis using the CS3 spectrophotometer facility. The CS3 spectrophotometer allows one to determine RBG components on selected sets of pixels. This we did on a cell by cell basis across the samples shown and others obtained. We then compared these to similar sample across the sepals and petals of the flowers as captured in situ.

We used several calibration techniques to obtain both absolute and relative calibration.

6.1.4 Experimental Results

The results of the analysis of the flowers were as follows. In Table 1 we depict three early Stout hybrids with both flower and cell. The flowers show that there is an abrupt change at the gross level in color. The cell data reinforces that fact. The cell either contains an anthocyanin for red or it does not. The change is an immediate and abrupt change in color.

The following Table summarizes several of the secondary pathway products and their colors. Note that the anthocyanidins present in a reddish manner, flavonols as cream and carotenoids as oranges. There are variations but we have shown how these can be normalized and measured, as well as separated when combined.

Class	*Agent*	*Color*[39]
Anthocyanidin		
	Pelargonidin	orange-red
	Cyanidin	purplish-red
	Delphinidin	bluish-purple
	Peonidin	rosy red
	Petunidin	purple
Flavonol		
	Kaempferol	ivory cream
	Quercetin	cream
	Myricetin	cream
	Luteolin	yellowish
	Agipenin	cream
Carotenoids		
	Carotene	orange
	Lycopene	orange-red

First, consider the Mikado flower. We collected a sample across the petal coloration from yellow to red. The sample is shown in the Table. The boundary between the two colors is seen at the cell level as an abrupt boundary, there is no slow degradation. Cells in the red region are clearly that there is a generation of peonidin or cyanidin in the red region. The specific identification of the specific anthocyanin or the complex may be determined by using the techniques detailed in McGarty, MIT, 2008, "Color". Namely one may use an inversion method to determine relative concentrations subject to appropriate normalizations. This can be achieved using RBG measurements alone.

The boundary effect is frankly a bit abrupt and it appears to be a validation of the Turing mechanism, namely that one sees that the pelargonidin pathway is either on or off. The spatial wave motion that one would anticipate with the Turing model is perceive in the macro analysis of the flower. If one looks closely at Mikado one can see the red and the flow to yellow at the edges. The anomaly is the yellow band along the central rib of the petal. One would suspect that there is a change in the diffusion coefficients at the cell level at this point. More detailed analysis of the cell structure must be performed to ascertain this.

[39] See Taiz p. 334 for the Anthocyanidin color and Bernhardt for the Flavonol and carotene.

We performed a more details analysis with Theron with samples taken from the throat region where there is a transition from red to the gold throated area. Again as can be seen in the Table the boundary is abrupt.

6.1.5 Discussion of Experimental Results

Patterns occur in Hemerocallis hybrids in a variety of manner and shape. Consider the following example of a 1941 hybrid produce by Stout, called Buckeye: We see what appears as an early eye pattern. This pattern had never appeared before. It shows a building of red, heavy on petals and light on sepals and it is bursting forth from the throat. It then ends abruptly. The question we pose is why? And how did this pattern result. It is patterns like these that we see a great deal of in the current generations of Hemerocallis hybridizing.

Petit, in the Daylily Journal of Summer 2007, described a multiple set of patterns that breeders were producing. He found the following to hold: The patterns take on a canonical set of forms. The forms are generally quite similar, perhaps because of common breeding practices or perhaps because of some underlying genetic makeup. These forms can be characterized canonical by a defining set of multiple overlays of tessellated secondary pathway expressions. Petit Patterns can therefore be explained and predicted. The patters described by Petit can be characterized as in the Table below.

Characteristic	***Turing Model***
Appliqué Throats	Unknown mechanism

Mascara Eyes or Bands	Demonstrates multiple layers of low spatial frequency outward growth of color.
Inward Streaks	If flower grows outward then the flow of control is unstable across new rows of growth.
Concentric Circles or Bands	If flower grows outward then the flow of control is unstable between new rows of growth.
Washed Eyezones	Ultra High intercellular instability, with almost localized oscillations allowing high spatial frequency of color change.
Stippling	High intercellular instability, with almost localized oscillations allowing high spatial frequency of color change.
Metallic Eyes	Unknown mechanism
Veining	Demonstrates multiple layers of low spatial frequency lateral growth of color.
Rainbow Edges and Midribs	Unknown mechanism

We can expand these descriptions if we create a collection of cells, say plant cells, and then color them to match the described and exampled patterns. We do this in the Table below. We also have shown in the Figure containing recent hybrids the actual physical embodiment of many of these patterns.

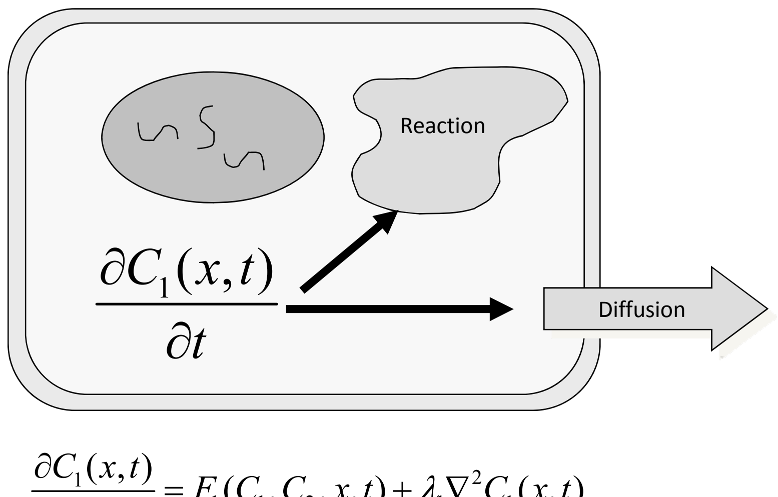
Mascara	***Inward Streaks***
Concentric Circles	***Washed Eyezones***
Stippling	***Veining***
Rainbow Edges	

What the Petit Patterns resemble is what we typically in oscillations occurring with the wave motion of second order spatial partial differential equations. By looking at the Petit Patterns as one with cells of a finite color contrast, constant hue, we can see that we can model this by means of such a second order equation but with nonlinearity in the output. That is we may have certain concentrations of enzymes but the dominant color could be controlled by the enzyme controlling the fastest or dominant secondary path. We now will explain this in the context of the linked genetic channel.

In this section we use the concepts of the Petit Patters and then combine them with the Turing model of morphogenesis to develop a verifiable and manipualatable system as regards to patters of the typ2 shown by Petit. We have shown that cells transmit to one another via the plasmodesmata. They communicate proteins and other concentrates from cell to cell. These thousands of small pipe ways create a diffusion process between the cells. We show a typical example in the figure below. The following is a detailed description of what Turing proposed in his model.

[Rate of change of concentration] = [Reaction Kinetics] + [Diffusion]

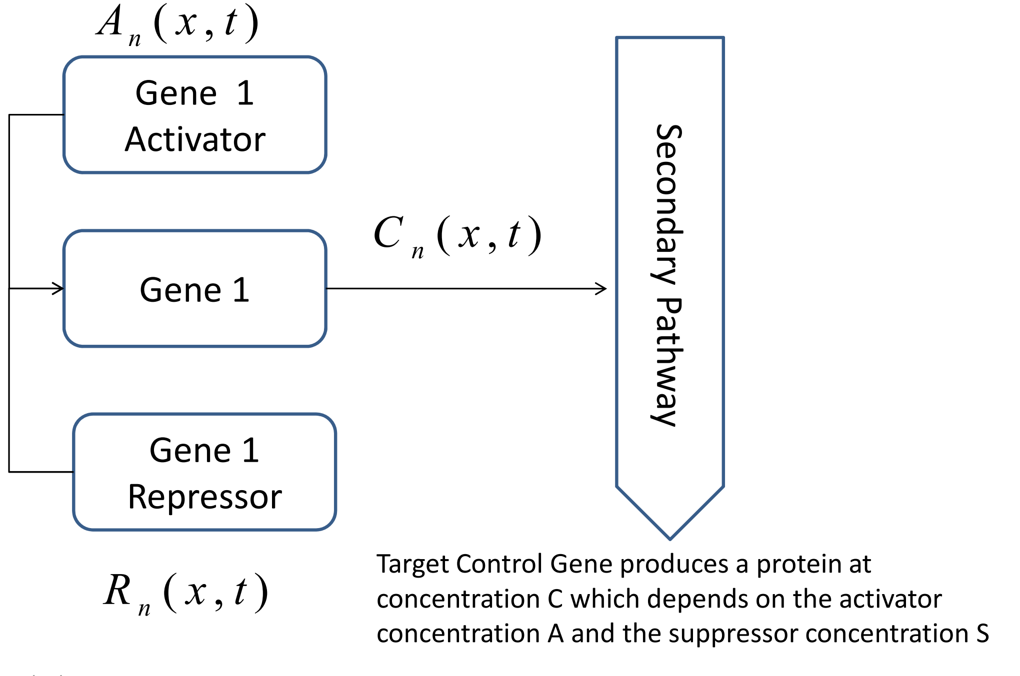
C_1 = concentration of protein or secondary c_1 in the cell



$$\frac{\partial C_1(x,t)}{\partial t} = F_1(C_1, C_2, x, t) + \lambda_1 \nabla^2 C_1(x,t)$$

6/8/2008 Tessellated Patterns in Hemerocallis 18

The following depicts the gene control paths that we will focus upon. The target gene produces a protein which enzymatically activates the secondary pathway producing the colorant. The target gene is activated by a gene which produces protein A and is repressed by a gene producing protein R. These proteins control the generation of C. In addition these proteins flow back and forth across the cell boundaries building up and decaying, as if in waves, and when the A exceed then we have activation and when R exceeds A we get repression.



The simplest model we have can then be stated as follows:

$$C_n(x,t) = \begin{cases} C_n(x,t) \text{ if } A_n(x,t) > S_n(x,t) \\ 0 \text{ if } A_n(x,t) < S_n(x,t) \end{cases}$$

$$\frac{\partial A_n(x,t)}{\partial t} = F_1(A_n, R_n, x, t) + \lambda_I \nabla^2 A_n(x,t)$$

$$\frac{\partial R_n(x,t)}{\partial t} = F_1(A_n, R_n, x, t) + \lambda_R \nabla^2 R_n(x,t)$$

The Turing Space is that space of a set of parameters, generally related to the enzyme (protein) reactions of the activator-suppressor genes which permit instabilities in the control mechanism of the Target Gene protein to the secondary pathway. Turing in 1952

showed that diffusion of the activator-suppressor proteins can cause instabilities, rather than the more common stable solutions. We now develop the following:

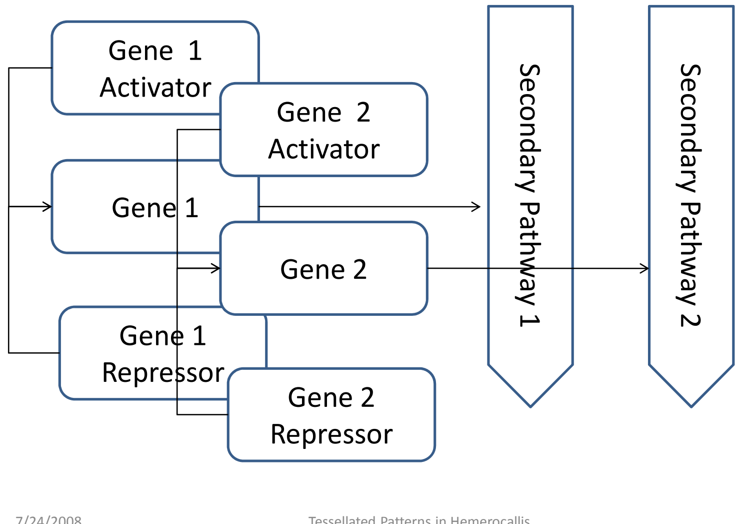
A model for the enzyme reactions in a competitive environment has been employed. A method to solve for the Turing space the diffusion model A model to apply the results to a single anthocyanin The ability to apply to multiple anthocyanin The ability to determine the analysis and the synthesis problem   The Turing model has been discussed earlier. What Turing proposed was that there was some chemical whose concentration made something one way or another. That this something then diffused throughout the organism in some manner and if it was greater in one part than a threshold the morphology was one way and if less the morphology was another. He had no underlying basis in the current understanding of genetics to put details to his models. We now have that detail.

We know that if we have an activator protein on a secondary pathway then that protein will cause the pathway to become active and create the secondary product, an anthocyanin. The more of that protein we have, the greater it concentration, the more secondary product we can get. This is P is the controlling protein concentration, we have:

$$\frac{\partial P_n(x,t)}{\partial t} = H(P_n, A_n, R_n, x, t) + D_P \nabla^2 P_n(x,t)$$

We show such pathways below. In the above pathway we have a complex but modellable set of interactions. They are characterized by:

When there are multiple A-R interactions then they add and the net result is an overlapping of the anthocyanin pathway products. The overlays can be shown to create the typical patters in the Petit list. The model allows for an analysis of any tessellated product and also provides a basis for determining what products are achievable as well as how to achieve them, at least at the genetic level.   Now we want to build on this model. First we must look at the dynamics of the activator and repressor genes and then we look at the dynamics of the controlling enzyme. Remember that the activator suppressor genes produce products which control the colorant gene. Let us now look at a single cell and look at the tempero-spatial dynamics of the concentrations of the products of the activator and repressor genes, A and S respectively. We assume we have a model as shown below:



Where in this model we have sets of genes and each has activators and repressors. Each gene may activate a separate pathway as we have shown. First we write the model for the controlling enzyme:

$$\frac{\partial P_n(x,t)}{\partial t} = H(P_n, A_n, R_n, x, t) + D_P \nabla^2 P_n(x,t)$$

In the above we show the concentration for the controlling enzyme in a cell for path n. It has a function H which results from a Michaelis-Menten pathway mechanism which we described earlier. From the Michaelis-Menten analysis before we have, if we assume some separate A, R process:

$$H(P,S,SP) = [+k_1 PS - (k_{-1} + k_2)C]Q(A,R)$$

where we had defined PS and C as before and where Q is a function of A and R which either turns on or off the process creating the P reactant. That is if A>R we have a reaction and otherwise we do not.

Thus P is also affected by concentrations of activator and repressor genes, A and R respectively, but in a binary manner. Second, now we write the general model for the activator and repressor product concentrations. As we have just discussed, the pathway activating protein is either on or off. If on we can then calculate its intensity and if off it is irrelevant. For the activator we have:

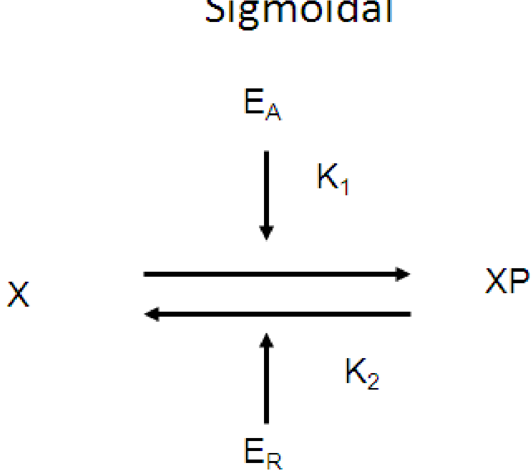
$$\frac{\partial A_n(x,t)}{\partial t} = F(A_n, R_n, x, t) + D_A \nabla^2 A_n(x,t)$$

and for the suppressor we have:

$$\frac{\partial R_n(x,t)}{\partial t} = G(A_n, R_n, x, t) + D_R \nabla^2 R_n(x,t)$$

Here we have A and R as the relative concentrations of the products of the Activator and Repressor genes. The F and G functions are the mass balance functions for this mix and the additional loss or gain come from the diffusion term. Here we assume that A and R may diffuse at different rates and this fact is key to the oscillations in space and in turn to the tessellation.

We now need a model for the interaction functions. We choose the model provided by Conrad and Tyson in Szallasi et al which is termed the phosphorlation-dephosphorlation model or the sigmoidal model[40]. We show its network below. Here we use the enzyme approach with one enzyme, the activator moving the production of the product enzyme and another the repressor enzyme driving the process backward. As we have done with the enzyme case we assume limited amounts and thus we have the
denominators in the equation.



[40] There are a multitude of models here.

This is the mathematical model we have deployed again using the same reference. This as we have said assumes that we have some form of enzymatic limiting reaction.

$$\frac{dx}{dt} = -\frac{k_1 e_A x}{K_{m1} + x} + \frac{k_2 e_S \left(x_T - x) \right)}{K_{m2} + x_T - x}$$

The above are also normalized concentrations. We rely upon the recent summary by Baici and the work of McMurray, Schnell as well as Szallasi and his co-authors. The generalized solutions we have are as follows, each normalized as we had done for the enzyme reaction. The equations are for the activator, repressor and product enzymes respectively.

$$\frac{\partial a}{\partial t} = f(a,r) + D_a \frac{\partial^2 a}{\partial x^2}$$

The same set of models for the repressor concentrations can now be developed except that we have a different diffusion constant.

$$\frac{\partial r}{\partial t} = g(a,r) + D_r \frac{\partial^2 r}{\partial x^2}$$

Finally for the controlling enzyme product:

$$\frac{\partial p}{\partial t} = h(p,a,r) + D_p \frac{\partial^2 p}{\partial x^2}$$

We can now linearize the system as follows about a point:

$$w(x,t) = \begin{bmatrix} a(x,t) - a_0(x,t) \\ s(x,t) - s_0(x.t) \\ p(x,t) - p_0(x,t) \end{bmatrix}$$

This will yield the following linear model:

$$\frac{\partial w}{\partial t} = Aw + K\nabla^2 w$$

where A is a 3x3 matrix and we have K also a 3x3 matrix. and where we have linearized the system to read as follows:

$$A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}=\begin{bmatrix} \frac{\partial f}{\partial a}..\frac{\partial f}{\partial r}..\frac{\partial f}{\partial p} \\ \frac{\partial g}{\partial a}..\frac{\partial g}{\partial r}..\frac{\partial g}{\partial p} \\ \frac{\partial h}{\partial a}..\frac{\partial h}{\partial r}..\frac{\partial h}{\partial p} \end{bmatrix}_{a_0,s_0} = \begin{bmatrix} f_a...f_r..f_p \\ g_a...g_r..g_p \\ h_a...h_r..h_p \end{bmatrix}$$

Murray shows that the following five properties are necessary and sufficient to determine the Turing Space for any reaction kinetics. These follow the stability requirements:

$$\frac{\partial f}{\partial a}+\frac{\partial g}{\partial s}<0 \quad \frac{\partial f}{\partial a}\frac{\partial g}{\partial s}-\frac{\partial f}{\partial s}\frac{\partial g}{\partial a}>0 \quad \gamma\frac{\partial f}{\partial a}+\frac{\partial g}{\partial s}>0 \quad \left(\delta\frac{\partial f}{\partial a}+\frac{\partial g}{\partial s}\right)-4\delta\left(\frac{\partial f}{\partial a}\frac{\partial g}{\partial s}-\frac{\partial f}{\partial s}\frac{\partial g}{\partial a}\right)>0$$

Now we can solve these equations and the concentrations for a and r are shown as follows:

$$a(x,t)=a(x)=\sum_{n=1}^{\infty}\eta_k\cos\left(2\pi\frac{x}{\kappa_a}\right)+\mu_k\sin\left(2\pi\frac{x}{\kappa_a}\right)$$

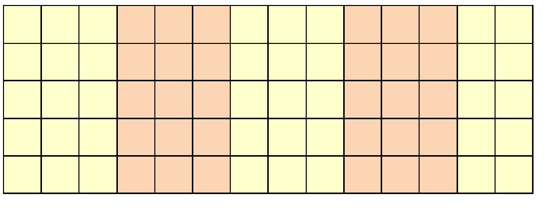
$$r(x,t)=r(x)=\sum_{n=1}^{\infty}\theta_k\cos\left(2\pi\frac{x}{\kappa_r}\right)+\xi_k\sin\left(2\pi\frac{x}{\kappa_r}\right)$$

$$d(x,t)=d(x)=a(x)-r(x)$$

$$D(x)=1\,\mathrm{sgn}(d(x))$$

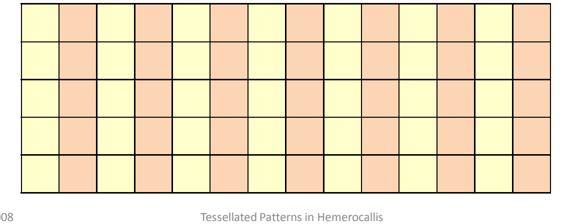
The k values of the wavelength are determined by the eigen values of the A matrix. What this model shows is that A and R has wavelike behavior for each anthocyanin dependent upon the diffusion coefficient for the specific proteins through the plasmodesmata. We now show several example of the solution to these equations. The following Table presents solutions determined via the approach in Murray. There have been many others over the years who have obtained similar results.

Example 1: p. 392 Murray, low spatial frequency instabilities across the cells showing wide striped variation. This assumed a single unstable secondary pathway which was on or off. It assumed also another stable secondary pathway.



7/24/2008 Tessellated Patterns in Hemerocallis 63

Example 2: p. 392 Murray, high spatial frequency instabilities across the cells showing wide striped variation. This also assumed a single unstable secondary pathway which was on or off. It assumed also another stable secondary pathway.



7/24/2008 Tessellated Patterns in Hemerocallis 64

We can now consider a simple model. This is a one dimensions, x axis only, model but it proves the point. Consider the following. We have two waves with the following amplitudes and wave numbers. Specifically we chose: k_A=0.1 and k_B = 0.2 and the coefficient of A to be 0.5 and the coefficient of R to be 0.2 The pattern is shown below. It shows a flip between pink and yellow. There is not direct relationship between these colors and any specific anthocyanin. This example can then be spread to two dimensions readily as is shown in Murray.

The above is a Pattern using 1 dimensional Turing Equation. Other more complex patterns can be readily generated. It should be remembered that in this pattern we assumed an activator and repressor gene and that if the concentration of one was greater than the other we generated one anthocyanin or the other. Thus the bi-color banding can be seen as above. In effect this is the coherence pattern seen in interference optics.

We may now consider two simple models for an analytical result. The first model is a linearized and uncoupled model and the second is linearized and couples. In both models we look solely at the concentrations of the A and R proteins.

Example 1: Linearized and Uncoupled Model, we assume that the reaction rate is dependent solely upon the active concentration and that there is a linear diffusion process. We let A(x) and R(x) be the concentrations of the activator and repressors as a function of distance. Thus we have:

$$\pi_A A(x) + D_A \frac{\partial^2 A(x)}{\partial x^2} = 0$$

$$or$$

$$\kappa_A A(x) + \frac{\partial^2 A(x)}{\partial x^2} = 0$$

$$\kappa_A = \frac{\lambda_A}{D_A}$$

$$and$$

$$\kappa_R R(x) + \frac{\partial^2 R(x)}{\partial x^2} = 0$$

This is the simple model for two waves of concentrations having a spatial frequency of the square root of the appropriate κ and that the waves of concentrations are independent, ascending and descending. Clearly they must be between zero and a maximum value.

Example 2: Linearized and Coupled, we now assume that the reactions, although linearized are coupled. Consider the following model,

$$\lambda_{A,A}A(x)+\lambda_{A,R}R(x)+D_A\frac{\partial^2 A(x)}{\partial x^2}=0$$
$$\lambda_{R,A}A(x)+\lambda_{R,R}R(x)+D_R\frac{\partial^2 R(x)}{\partial x^2}=0$$
$$or$$
$$\kappa_{A,A}A(x)+\kappa_{A,R}R(x)+\frac{\partial^2 A(x)}{\partial x^2}=0$$
$$\kappa_{R,A}A(x)+\kappa_{R,R}R(x)+\frac{\partial^2 R(x)}{\partial x^2}=0$$

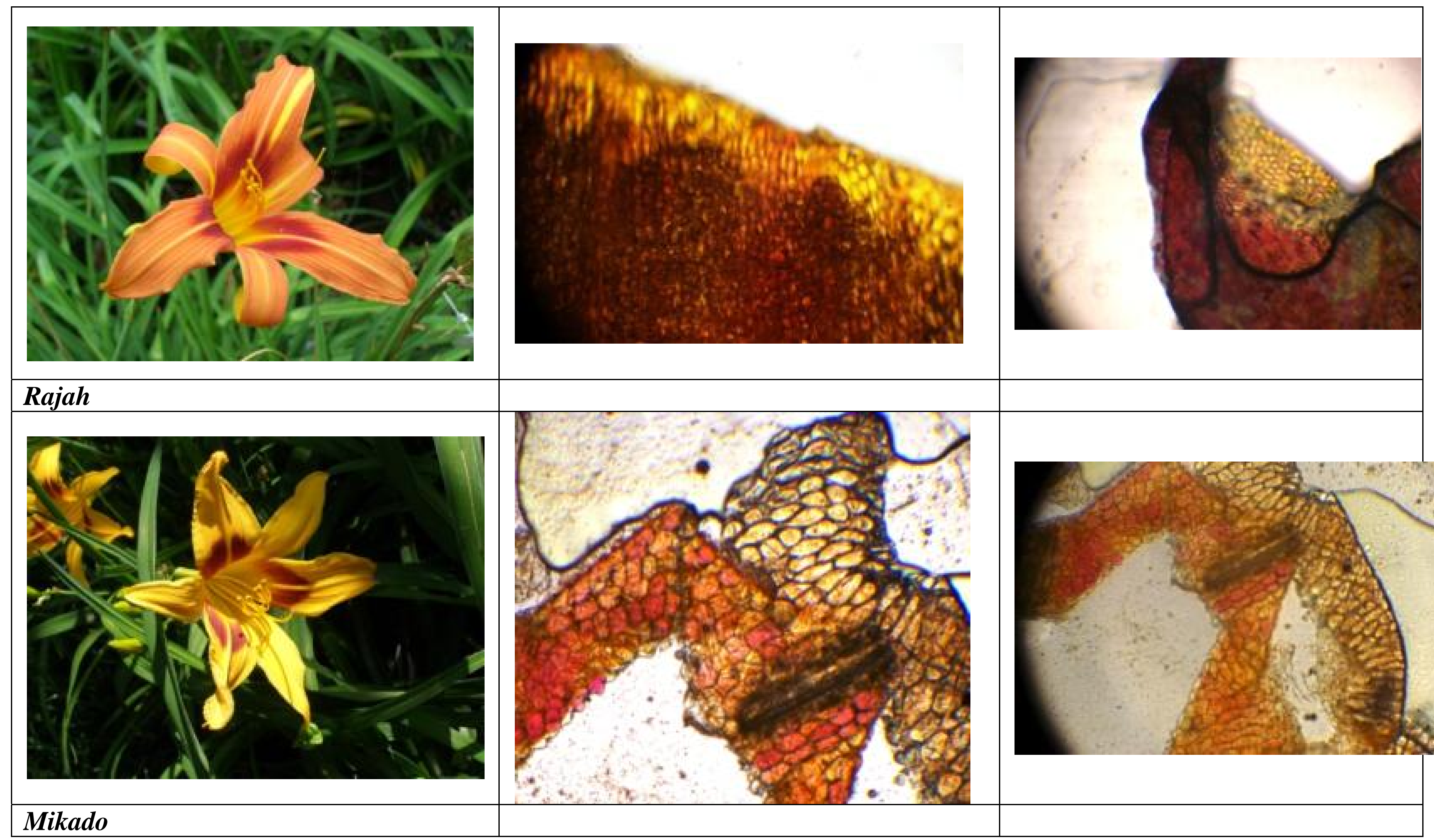
Now this is a bit more complicated that the first example and it can be solved analytically but we can make a few simplifications and see what the impact of some limited coupling would be. We can now write the above as follows:

$$\begin{bmatrix}\frac{\partial A}{\partial x}\\ \frac{\partial^2 A}{\partial x^2}\\ \frac{\partial R}{\partial x}\\ \frac{\partial^2 R}{\partial x^2}\end{bmatrix}=\begin{bmatrix}0.....1.....0.....0\\ \kappa_{A,A}.0.....\kappa_{A,B}.0\\ 0.....0.....0.....1\\ \kappa_{B,A}.0.....\kappa_{B,B}.0\end{bmatrix}\begin{bmatrix}A\\ \frac{\partial A}{\partial x}\\ R\\ \frac{\partial R}{\partial x}\end{bmatrix}\backslash$$

This is readily solvable and depending on the relative values one may see that the coupling merely perturbs the spatial frequency obtained from the first example.

From an experimental perspective two final observations may be made' First, one may determine the constants of reaction and diffusion by examining the pattering of the specific flower. We have demonstrated that elsewhere (see McGarty, MIT, 2008). Second, one may reverse the problem and seek to affect a specific pattern. That pattern will then stipulate a set of diffusion equations and in turn a set of constants. These can now be

genetically engineered (again see McGarty, MIT, 2008). Thus the analysis and models developed herein take what Turing had developed more than fifty years ago, and apply specifics for the patterning of flowers.



Rajah

Mikado

Theron		

EXAMPLES OF HYBRID PATTERNING

6.2 Gene Expression in Plants: Use of System Identification for Control of Color

We now address the intra-cell issue of color. Plant color is controlled by the expression of certain chemicals in secondary pathways in plant metabolism, especially the anthocyanin pathways. The secondary pathways are controlled by protein concentration resulting from selective genes and their rate of expression. This paper analyzes these complex pathways as state machines and uses various system identification techniques to identify the mechanism of the pathways. Using the mechanism of the pathways provides a mechanism for control of the pathways to produce a desired color

The genetic control of the expression of color in plant flowers is currently understood to be effected at several levels in the genetic and chemical pathways in plant cells. The most recent understanding is that there are secondary chemical pathways which produce such chemical products as anthocyanins and these pathways are controlled by enzymes, proteins, which are the products of genes in the plant cell. The plant flower color is then a result of the concentrations of the anthocyanins generated in these secondary pathways, for example. Each secondary pathway is rate controlled and is driven by enzyme concentrations. The greater the concentration of the enzyme, the greater the specific anthocyanin concentration. At the gene level other controlling factors as well. There are other secondary genes which may activate or suppress the gene which generates the activating enzyme. The combination of these elements can be expressed as a dynamic system and the process of determining the characteristics of this dynamic system can be posed as a classic system identification problem. That problem is analyzed in this paper and we focus on a specific species, Hemerocallis of the family Liliaceae, a monocot plant found in China, Korea and Japan. The inverse of this problem is also considered, namely if we desire a specific color, then what we seek to control at the gene level to effect this desired color. This is the controllability problem.

The following questions are addressed in this analysis and model development:

1. Given a dozen or more species of plants which are relatively stable and consistent in the wild, how does the variation in color in hybrids arise? What is the cellular basis of color, and what is the genetic set of mechanisms which controls it?
2. Given the complexity of color, form and variegation in the hybrids, what is the genetic basis for the intracell and intercell control mechanisms? For example, how

are such colorations as eyezones formed and what is the intercellular communications mechanisms which effect this?
3. Given what now appears to be a set of well-understood pathway- control mechanisms by enzymes produced within the cell and the gene control mechanisms for expression of these proteins, how are these combined to produce intra cellular coloration and what are the inter cellular communications which spread the colors out over the inflorescence?
4. Given that we can answer the above, can we generate a mathematical system for gene expression and control and using the model to solve the coloration problem using system identification or inversion?
5. Given that we could solve the above problem, then how could we apply positive control to coloration and produce whatever color we desire?

Our approach in this paper is straightforward. We focus on a specific genus, Hemerocalis, and on a specific part of the plant in that genus, the inflorescence.

There are two areas which are developed herein. They are the characterization of flower color and the system structure of genetic control of secondary pathways.

Flower Color: We present an overview of the process of developing color in flowers. We present an overview of the anthocyanins, flavonols, and carotenoids. We review their pathways and summarize recent research which had identified the enzymes on each link of the pathway and the genes controlling those enzymes. This has been accomplished over the past few years and is critical to the understanding of the overall system approach.

Cell Genetics: We provide a detailed overview of cell genetics and how activators and repressors are key elements in the overall expression of enzymes and in turn the development of color. We review the cell elements and especially the process of gene expression. We discuss activators and repressors and the mechanisms of their actions. Their existence results from the work of Monod and Jacob in the early.

6.2.1 System Models for Gene Expression:

Recently the biological community has applied system models to biological systems. We build on that effort and develop modals for the expression of flower colors. We recognize that color is a result of a mixture of secondary plant products such as anthocyanins. We can from the color of a flower determine what the mixture of each anthocyanin is. The concentration of an anthocyanin is a result of the concentration of the enzymes in the pathway which produces the anthocyanin, and typically the lowest enzyme concentration is the dominant factor. We also know that the concentrations of the enzymes is a result of activators and repressors, proteins also generated in a cell, which turn on or turn off the enzyme controlling the pathway. The work of Winkel

Shirley provides an excellent review of the status of systems techniques currently employed in the genetic analysis of organisms.

Combining these ideas we can develop a top down system model for color. The output or observation equation is the color, and the system equation is a dynamic process wherein the states are the protein concentrations from a large enough set of gene expressions, wherein genes are allowed to control other genes via an nth order dynamic process. We also allow for uncertainty by adding a “noise” process which converts the overall system model into a linear dynamic stochastic system with observables. We extend that model from a single cell to a matrix of interconnected cells. This allows us to explore the processes one sees in the development of eyezones and other sharp transitions of color in flowers. We use models which have been previously studied for color variation and apply those to the flower. In particular we will focus on each of the biochemical elements shown below.

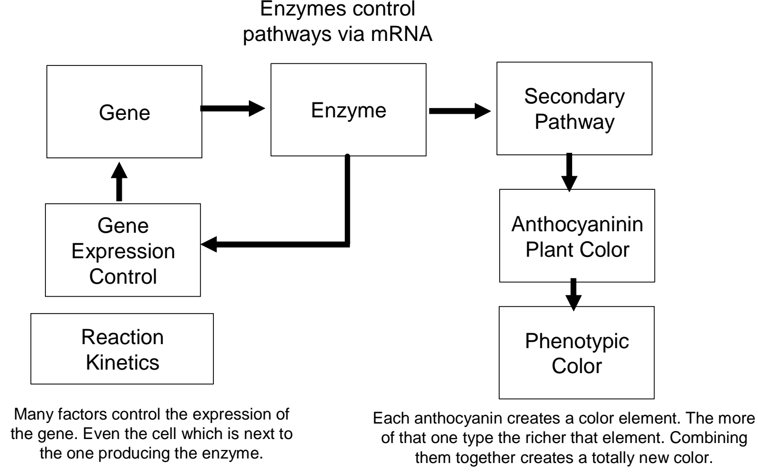


Figure 1: Basic Model for the production of secondary pathway agents controlled by underlying gene expression.

6.2.2 Prior Work

The key prior works fall into three categories; (i) underlying genetic studies and understandings of the genus, (ii) detailed elucidation of the control of pathways and the effecting gene sequences, and (iii) the development and application of models for the analysis and synthesis of gene expression.

6.2.2.1 Genetic Structure of Hemerocalis

Various recent works have provided detailed genetic analyses of the genus. Specifically has provided a detailed study of three populations of the species hakuensis and has shown that there is a significant intra-species variation. This has been known for many years. This was a problem for many plant sytematicists who had few examples of species available and used this limited number to describe the species. It is necessary to perform extensive field work to fully understand the intra-species variation.

Specifically have studied the species citrina. Their work included a detailed analysis of certain exons and an understanding of the evolution of this species. They have begun to establish a basis for genetic analysis of within species characteristics. Extensive work in have shown a detailed analysis of the full genetic variation in the genus using AFLP markers. They have also extended this to include many of the current common hybrids. Their key observation is that in the recent hybrids that they studied the genetic similarity has increased by approximately 10%. This demonstrates the rather interesting effect that if the genetic diversity is decreasing and the phenotypic changes are increasing then it clearly must be via expression.

6.2.2.2 Gene Expression and Pathway Control

Within the last five to ten years there has been considerable growth in the understanding of the control of the pathways which provide for color. The recent work in provides a superb summary. The author reviews prior efforts and puts the entire pathway management into perspective. The work in provides all controlling proteins and their causing genes, and does this for anthocyanins and flavones and isoflavonoids. The relationship to the abundance of effecting enzymes and anthocyanin expression as well as flavone expression is clearly demonstrated. This gives us a key in the development for our model.

The work Mol predates Winkel Shirley and is the seminal paper on genetic control of pathways. Mol and his colleagues have continued to publish their results on further clarification of the pathway management by gene expression. Holton and Cornish was the first to publish the full pathway and that work is a major contribution to the field. Work on carotenoid pathways has been completed Naik and then Bartel.

6.2.2.3 Modeling of Gene Expression: Analysis and Synthesis

The development of systems modeling for gene expression has most recently been exceptionally well articulated in Szallasi. In this work a collection of authors who are a mixture of systems experts and micro biologists present an up to date summary of all key works in this area. The work Hatzimanikatis is also an excellent modeling tool which applies a more Boolean approach to modeling the expression of genes. However the authors also extend their models to include mRNA and other pathways in a linear time varying system model as well. This latter approach coincides with the recent directions portrayed in Szallasi and is consistent with the approach taken herein. In Vohradsky is

provided a neural network approach to the understanding of gene expression. Although highly flexible this model is at best amenable for limited simulation analysis. Work in Perkins has provided a detailed systems model for expression using classic dynamic systems models. Chen has also provided a detailed dynamic model using their "differential equations" approach. We see that Perkins and Chen have a great many similarities, as does the collection of authors in Szallasi but they all seem at best to be just becoming aware of the wealth of well understood theory in the control and estimation area in McGarty.

6.2.3 Flower Color Expression

There is an extraordinary variation in the color of the hybridized flowers of the genus Hemerocallis. In a little over a hundred years hybridizers have taken the dozen or so species, all predominantly yellow, orange or red, intermixed them and as a result have created a very complex set of flowers with characteristics which differ dramatically from each of the species. The species have managed to maintain their separate identities over thousands of years but in a small fraction of time we have been able to introduce multiple forms and colors. To understand this process we first have to understand where the colors come from. How do we get purple from a plant which is red, yellow, orange and possibly even brown? How are the colors made? In this paper we focus on inflorescences of one color. The issue of variegated inflorescence has been studied initially in 1949 [29] in a brilliant paper before the Watson Crick model was developed and his analysis is expounded upon in Murray. The Turing model is more complex than what we present here and will be detailed in previous section.

The first step in understanding that process is to understand the pathways that lead to color production in a single cell. Then we can address the issue of multiple cells and finally how the cells communicate. For example, how do we get an eyezone? Why if a cell is white do we go so abruptly to a purple eyezone? What is the mechanism for this process? We begin the exploration of this issue with a analysis of the underlying pathways.

6.2.3.1 Pathways and Enzymes

Pathways are nothing more than a set of chemical reactions which get us from some primitive chemical to a more complex but useful chemical structure. In fact the pathways may be just a set of processes going from any one chemical structure to another independent of the nature of the starting and starting chemical. Some pathways are linear going from a beginning to an end and some are circular taking us from the beginning and back again (the Krebs cycle is an example). What makes the pathway work? Just three elements are required: (i) the underlying chemical constituents, (ii) some form of energy, (iii) generally some form of facilitation such as a catalyst and in our analyses this is an enzyme. We have the pathway but it is facilitated by an enzyme, a protein. The protein is generated by a gene. And the gene is activated

by some other element, generally another protein. In our case shown below the output is some anthocyanin. The more of the enzyme, the more the gene expresses itself and the more anthocyanin we get; this is the basis of enzyme reaction kinetics. Thus if we can get the gene to express then we get more of that specific anthocyanin, more pelargonidin for example. We defer to the next section how we get this gene to express so strongly. The opposite is also true: if we can suppress the gene then we can get less and even possibly no anthocyanin from the pathway. This is the first step in the development of an overall system model.

6.2.3.2 Anthocyanins

Let us consider our first pathway. This is the pathway which creates. The anthocyanin molecule is shown below. Note on the B ring we have six sites to which we can attach differing molecular chains. This will be an important element when we see the different configurations and their implications.

The anthocyanin or anthocyanidin molecule comes from two different pathways. One is from the shikimic pathway and the other from the malonate pathway. This means that we have to understand both pathways to understand the ultimate abundance of the product.

In the Table 1 below we list the anthocyanidin and its resulting color. Each is obviously named after their related flower and each results an anthocyanin of a different color.

TABLE 1: ANTHOCYANIN AND ASSOCIATED COLOR RESULTING FROM ITS ACTIVATION

Anthocyanidin	*Color*
Pelargonadin	orange-red
Cyanidin	purplish red
Delphinidin	bluish purple
Peonidin	rosy red
Petunidin	purple

Each of the colors is the weighting of a red, green and blue combination which best matches the color. Thus one can, in an 8 bit color scheme for example, as one would find in any computer color scheme, get the resulting anthocyanin colors by blending the R, B, G elements to yield what we seek. This relating the colors back to RBG is critical since it gets reflected in the ultimate flower color.

Now if we assume we have only anthocyanins for color, and that we have the above combinations available, we ask how we combine these colors in a weighted manner to obtain the desired color. This approach is critical to our overall understanding. First we show by a weighted RBG we get the color we seek or the color which is presented. Then we assume we can do the same for each anthocyanin, and thus we can create any desired color from a weighted collection of anthocyanins. This means that we can determine what the relative percents of expression of any anthocyanin are and this lets us go back to how strongly the gene for that anthocyanin is expressed. The model we presented earlier will be a key element in this overall process.

Now let us start with a simple expression. This text presents a detailed analysis of how color is characterized in the red/green/blue model. This model carries over directly to the computer color model that is currently in use. We have used it as a core baseline for the observable from the overall secondary pathway process. Thus for any color we can write:

$$\text{Color}=w_1 \langle \text{Red} \rangle + w_2 \langle \text{Blue} \rangle + w_3 \langle \text{Green} \rangle$$

$$\begin{aligned} &\langle \text{Red} \rangle = \text{color base of Red} \\ &\langle \text{Green} \rangle = \text{color base of Green} \\ &\langle \text{Blue} \rangle = \text{color base of Blue} \end{aligned} \qquad (1)$$

and if we define:

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

$$s = \begin{bmatrix} \langle \text{Red} \rangle \\ \langle \text{Green} \rangle \\ \langle \text{Blue} \rangle \end{bmatrix} \qquad (2)$$

resulting in:

$$Color = w^T s \qquad (3)$$

Likewise we could state this by means of some combination of anthocyanins and their related colors. Now we can define any color as a combination of the three anthocyanin concentrations and the concentrations effects on color by using the basic red, green and blue bases as below:

$$Color=m_1\langle Pelargonidin\rangle+m_2\langle Delphinidin\rangle+m_3\langle Cyanidin\rangle$$

$$\begin{aligned}\langle Pelargonidin\rangle&=g_{1,P}\langle Red\rangle+g_{1,P}\langle Blue\rangle+g_{1,P}\langle Green\rangle\\ \langle Delphinidin\rangle&=g_{2,D}\langle Red\rangle+g_{2,D}\langle Blue\rangle+g_{2,D}\langle Green\rangle\\ \langle Cyanidin\rangle&=g_{3,C}\langle Red\rangle+g_{3,C}\langle Blue\rangle+g_{3,C}\langle Green\rangle\end{aligned} \quad (4)$$

We can measure the coefficients g in the above using standard colorimetry. If we define a matrix G as follows:

$$G=\begin{bmatrix} g_{1,P} & g_{2,P} & g_{3,P} \\ g_{1,D} & g_{2,D} & g_{3,D} \\ g_{1,C} & g_{2,C} & g_{3,C} \end{bmatrix} \quad (5)$$

Then we can determine the color as a simple product and the m values can be determined directly.

$$Color=m^T Gs=w^T s$$

$$(6)$$

$$m=G^{-1}w$$

The above analysis shows us that we can analytically determine the expression of the anthocyanins from the color of the cell by means of the above formulas. Color is determined in the Red, Blue, Green approach by weighting each of this prime colors by some weight w. Also we can obtain the same color by weighting the alternative colors as associated with the anthocyanins present by a similar weight in this case m and m is directly related to the concentration of that anthocyanin. These are relative expressions but by benchmarking any one element we can make them all absolute in the cell as well.

6.2.4 Other Color Elements

Anthocyanins are not the only elements which are secondary products which produce color. There are three classes of chemicals which give rise to color; anthocyanins, flavones or flavonols, and carotenoids. The Table 2 below depicts the different elements and their colors. The approach we took above for the anthocyanins can be used for the

flavones and carotenoids as well. It should be noted that there may not be a unique solution here but there are several possible. The solutions can be narrowed down by actual determination of one to three elements as baseline. The other two general classes are the carotenoids, the orange colors, and the flavonoids, the more white type of colors. We summarize the colors in the following Table 2.

TABLE 2: SUMMARY OF THE THREE MAJOR CLASSES AND THEIR AGENTS AND RESULTING COLORS.

Class	*Agent*	*Color*
Anthocyanidin		
	Pelargonidin	orange-red
	Cyanidin	purplish-red
	Delphinidin	bluish-purple
	Peonidin	rosy red
	Petunidin	purple
	Malvinidin	
Flavonol		
	Kaempferol	ivory cream
	Quercetin	cream
	Myricetin	cream
	Isorhamnetin	
	Larycitrin	
	Syringetin	
	Luteolin	yellowish
	Agipenin	Cream
Carotenoids		
	Carotene	orange
	Lycopene	Orange-red

6.2.5 PATHWAYS

In this section we present the pathways for the three classes we have described above. We first present an overview of the pathway and then we present the details of the pathway and the enzymes used in each step. The key observation is that we must have enzymes to go from step to step in the pathways and that if any one enzyme is missing we cannot proceed on that path, and further the path with the small amount of enzyme becomes the limiting path. Thus, we do not have a one to one map here. The production of any one anthocyanin, for example, if limited by the lowest produced enzyme, and the other enzymes may be present in abundance.

Figure 2 shows the overall pathway for all elements.

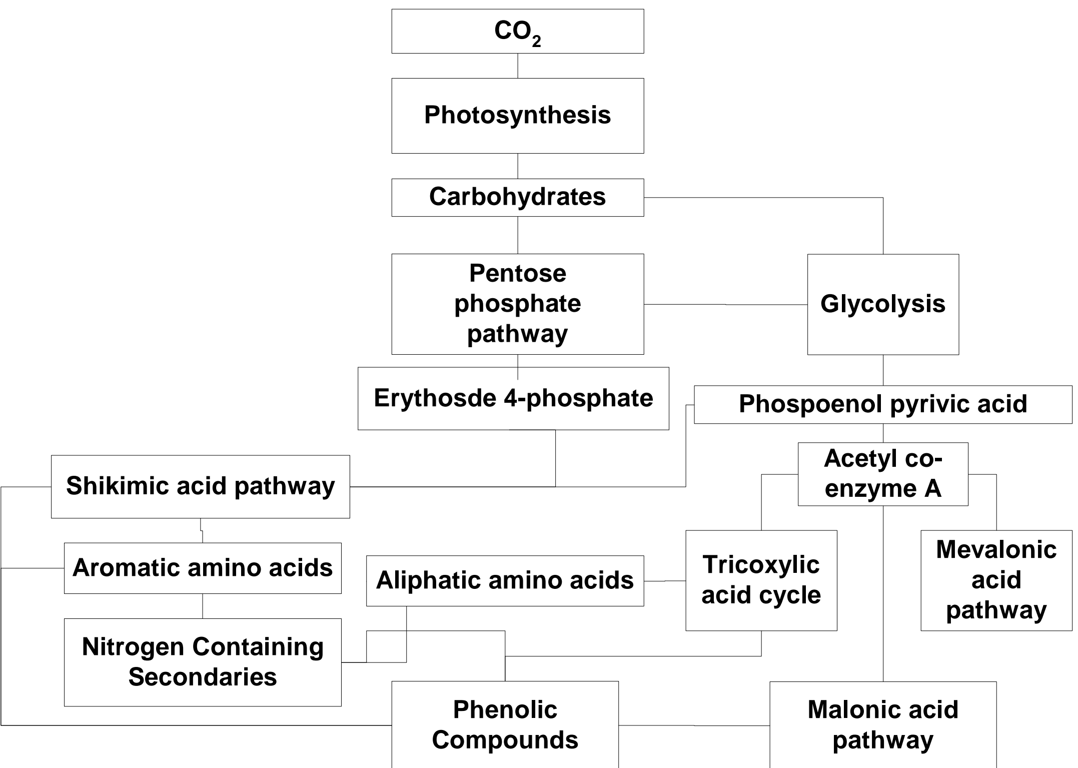


Figure 2: Pathway for the production of anthocyanin from Carbon Dioxide.

The above shows how we start from CO_2 and then go through a variety of other pathways. We will review those pathways in some detail since it is the enzyme control in them which is key.

The pathway moves forward at a reaction rate which is determined by the concentration of the reaction supporting enzymes. This is modeled by standard enzyme reaction kinetics as is provided in [32], [43], and [10].

6.2.5.1 Anthocyanin Pathway

The anthocyanin pathway with the controlling enzymes is shown in Figure 4. The enzymes are presented in the arrows linking each step in this pathway. This pathway shows the start as a sugar element and then goes to phenyanaline and then down through the chain to one of the four indicated anthocyanins.

FIGURE 4: SPECIFIC ANTHOCYANIN PATHWAY FOR SPECIFIC PRODUCTS.

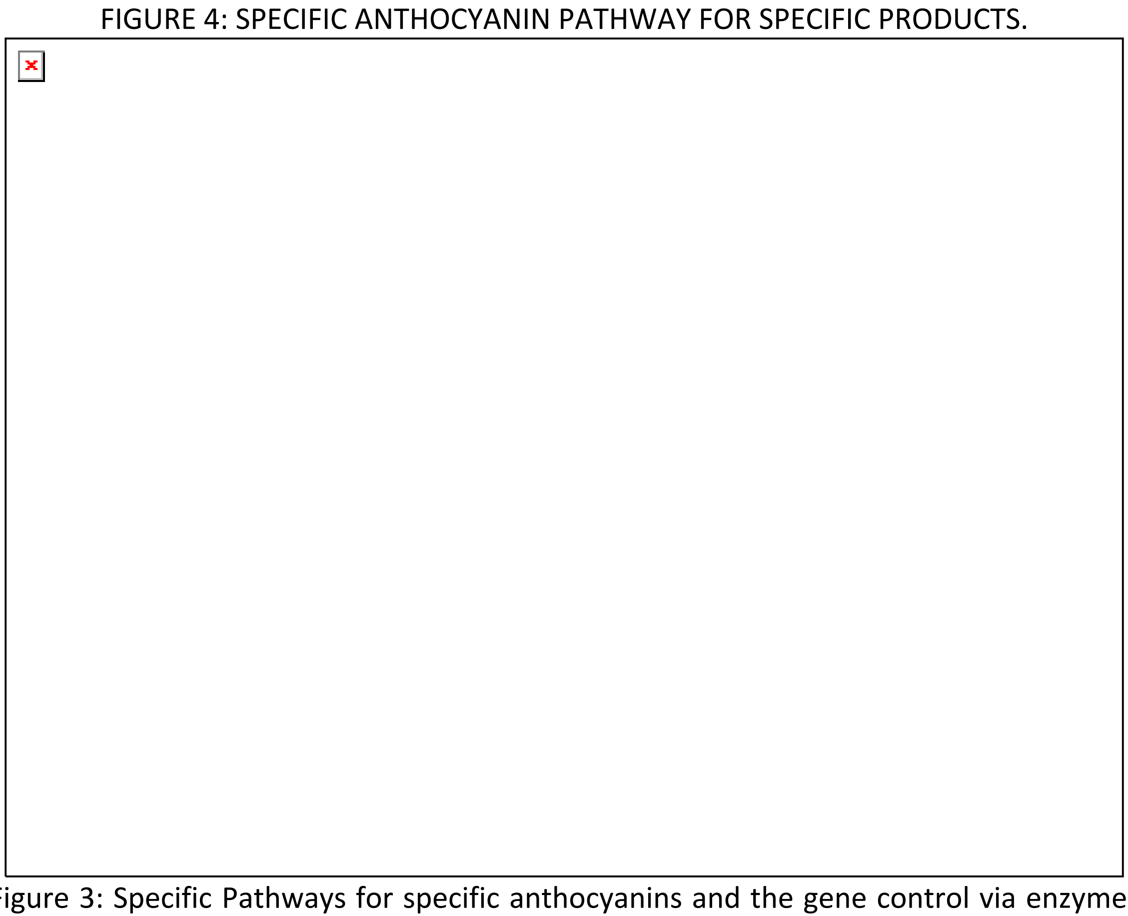


Figure 3: Specific Pathways for specific anthocyanins and the gene control via enzymes in specific pathway transitions.

Note that at each step there is an enzyme element. The genetic loci for cloned flavonoid enzymes in Arabidopsis are shown in the following Table 312.F

TABLE 3: ACTIVATING ENZYMES AND THE ASSOCIATED CHROMOSOME AND GENE LOCATION

Enzyme	Locus	Chromosome	Map Position
CHS	tt4	5	7,050 kb (MAC12)
CHI	tt5	3	21,000 kb (T15C9)
F3H	tt6	3	19,600 kb (F24M12)
F39H	tt7	5	4,400 kb (F13G24)
FLS	fls1<Enc	5	FLS1: 4,700 kb (MAH20) FLS2-5;: 32,150 kb (MBK5) FLS6: 24,350 kb (MRH10)
DFR	tt3	5	23,800 kb (MJB21)
LDOX	tt19	4	16,900 kb (F7H19)
LCR	ban,ast d	1	26,800 kb (T13M11)

What these processes point out can be summarized as follows:

1. There are common pathways which are operational in all plants for the generation of the pigments.

2. Enzymes used as activators modulate the amount of production of the enzymes.

3. The products of these pathways, the anthocyanins, are driven by the concentration of the facilitating enzymes.

Secondary products always have this type of production process. As we look at a cell, from a system point of view we see facilitating proteins and secondary products. The concentrations of the secondaries are proportional, in some general way, to the concentration of the facilitating proteins. However we see there are many facilitating proteins which may make this a more complex analysis.

6.2.5.2 Carotenoid Pathway

The carotenoid pathway is shown in [30]. It is similar in many ways to the anthocyanin.

6.2.5.3 Flavonol Pathway

The flavonol pathway is identical to that of the anthocyanin and is detailed in the work of [17].

6.2.6 Expression Analysis and Implications

In this section we develop a systems approach to the problem of color analysis and synthesis. This work is based upon the work in McGarty (1971) which focused a systems approach to the overall identification problem.

6.2.6.1 Approach: Engineering versus Science

The approach we take in this paper is an engineering approach rather than a biological approach. Our interest is in developing a model or sets of models which allow us, by a verifiable means, to show how the genes react and interact to produce the plant colors. We can compare this to the engineering approach to circuit design of transistor circuits versus the science of understanding the semiconductor from the point of view of detailed quantum mechanical models. The biologist in our approach is akin to the physicists. Our approach is akin to the engineer knowing that there is some set of physical processes inside the semiconductor which may clearly be important to the physicist but the engineer's interest is in designing and analyzing the circuit element to produce a system that behaves in a specific way.

Thus for an engineer, if we increase a current here we get a decrease or an increase at some other point. The engineer creates a world view of a macro set of processes and models the details of the biologists in our case with a few set of equations which show the results of increases and decreases. One must be able to make measurements to show that the processes predicted indeed occur, to a reasonable degree of accuracy. Then one can analyze a genetic circuit and then in addition one can design a genetic circuit. We then can understand where the colors come from and possibly engineer the genes to develop and deliver on colors we desire.

6.2.6.2 A Control Paradigm

The expression regulator for any gene may be an activator or suppressor gene. It may be a result of a gene expression in the cell itself or quite possibly as we shall discuss fed through from another cell. There are many of these regulatory cycles and they are all interconnected. This basic paradigm is one of hundreds or thousands of such interconnected flows.

In developing our models we will use this construct. However, we can frequently focus on natural clusters of related genes. They may be a dozen or more such related genes in each cluster and possibly hundred of such clusters. Although cells and their proteins may affect all other cells, only a few of the genes regulated have a significant level of regulation. The low levels of "regulation" we shall consider just as noise.

6.2.6.3 A Model for Secondary Production

A system model for the relationship between the genes, proteins and secondary path chemicals can be developed. We assume we know or can determine the following:

1. The secondary pathway chemical steps are known. This includes what enzymes modulate what transitions in the pathway.
2. The resulting concentrations of the products from the secondary pathways are proportional to the concentrations of the enzymes acting as catalysts on the pathways. The pathways follow known enzyme reaction kinetics.
3. The concentrations that result from the secondary pathways are reflective in the phenotype characteristic perceived.
4. The phenotype characteristics perceived are measurable and can be analytically related to concentrations in secondary pathway products.
5. The genes which produce enzymes which modulate secondary pathways are known in detail.
6. That activator and repressor genes of the modulator genes are known or knowable. Their specific effect on the modulating gene does not have to be known a priori.
7. There exists a database of genes which are modulator, repressor, and activator characteristic from which one may be able to analyze their levels of expression using microarray or similar technologies.

In the event that these assumptions are valid, which is the case now for many plants, as well as many animal models including humans, then one can develop models to determine the "system model" of the genes and the secondary pathway elements. This is the "analysis" portion of the system. It provides the elements of the system and quantitative values for its dynamic behavior. The second portion is the "synthesis" portion. In the synthesis portion we assume we have determined the values to model the dynamics of the system. Next we seek to drive the system to a desired state, in this case a desired flower color. To many this is the "blue rose" or the "blue daylily" issue. We commence with a model for the gene and its control, the secondary pathway and complete the analysis issue. The system model is depicted in the following Figure 4.

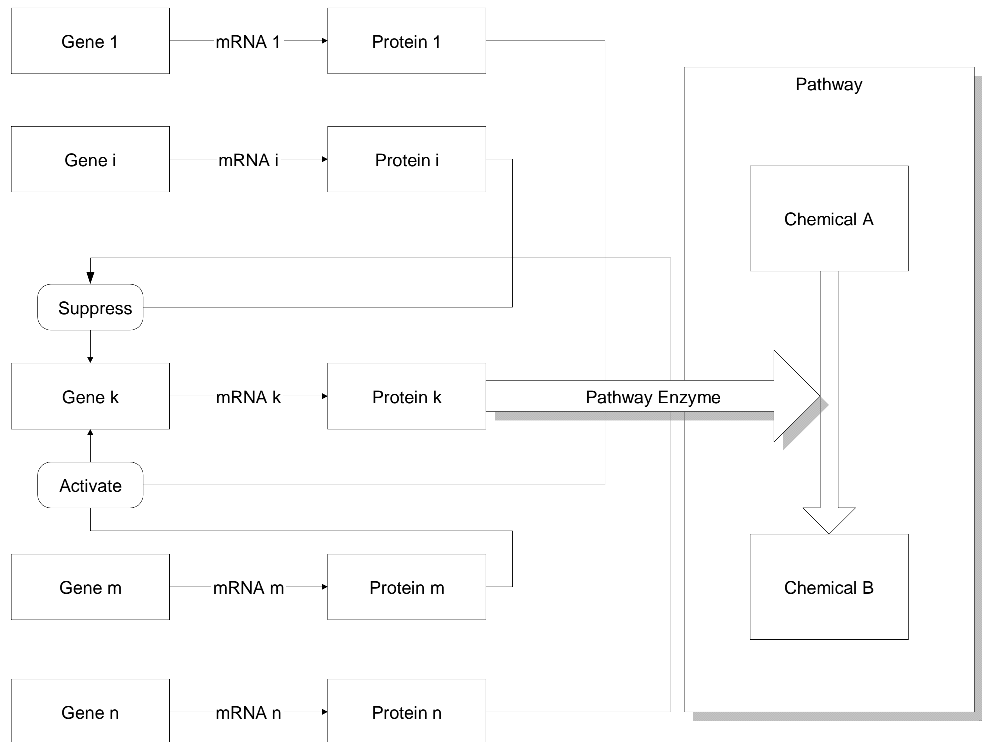


Figure 4: Model of pathway control and linkages between activating and suppressing genes.

The Equation (7) is an example of a general form for the expression of genes by means of the concentrations of the proteins related to specific genes. In the equation the expression of n genes is given as an n X 1 vector and the time variation of that vector is expressed as the sum of three other vectors; one is a vector which is dependent in a nonlinear manner on the concentrations themselves as well as time, one which is a function of some external set of influences which is a q x 1 vector u as well as time and an n x 1 vector which acts like random noise which accounts for a combination of noise and uncertainties.

$$\frac{dx}{dt}=f(x,t)+g(u,t)+n(t)$$

$$x_k=\left[\text{Concentration } P_k\right] \qquad (7)$$

A system model for the secondary pathways may also be developed. In this embodiment the following equation is an example of a general form for the secondary pathway. The m x 1 vector z represents what is observed and what in turn produces the phenotype characteristics. In the case of flower color the z elements may be the anthocyanin concentrations for example. In the equation below the z vector is a

nonlinear function of the protein or enzyme concentrations plus added random noise which represents uncertainties as well as natural external disturbing measurement phenomenon, where we have used w(t) a white noise process to account for both measurement model inaccuracies and measurement errors.

$$z(t)=h(x(t),t)+w(t) \quad (8)$$

The general models used above in this embodiment may be simplified by using standard techniques of linearizing them. The result of such a standard linearization process is shown in the following equation (9). In the equation below the x vector is the protein concentrations resulting from genetic pathways and the z vector is the concentrations of the anthocyanins or in general they are the concentrations resulting from the secondary path chemical products. In this model the relationship of the protein concentrations to secondary chemical concentrations is assumed known or knowable. In this embodiment the elements depicting the dynamics of the protein concentrations are assumed unknown but can be ascertained quantitatively by means of the methods discussed herein.

$$\frac{dx}{dt}=Ax+\sum_{i=1}^{N} g_i p^T D_j x+o(x)+g(u)+v$$

$$z=h(x)+w=Cx+w$$

$$z=\begin{bmatrix}[\text{Pelargonidin}]\\ [\text{Delphinidin}]\\ [\text{Cyanidin}]\end{bmatrix} \quad (9)$$

The above model may be further linearized to yield a simple linear system model which can be used in this embodiment. This simple linear model is shown in the following set of equations. The vector x is the set of protein concentrations resulting from the set of n gene expression interactions. The matrix a depicts the interaction between all of the genes as activators or repressors. the vector u is a known or unknown independent driving vector to the gene expression product. In this embodiments the matrix A will be determined by means of the procedures provided in this embodiment. The second equation depicts the steady state solution of the first equation. It should be noted that the steady state model is acceptable for plant color but the dynamic model may be required in many other systems where there is a dynamic portion to the system such as when in a human hormone release is involved. For the most part, however,

steady state is adequate.

The steady state solution depicted below use the vector u and the inverse of the matrix A. In the equation below the vector x is composed of protein concentrations which effect secondary pathways such as the ones for the anthocyanins as well as genes which activate or repress the genes which directly express for color as shown in the equation (10) below. In (11) the concentration of secondary chemical products, z, are shown to relate to the concentrations of the proteins resulting from the first process. The relationship is via a matrix C which is in (12). Finally in the equations below the weight elements of the color expression, the vector elements denoted by m, are shown to be obtained from the concentrations of the secondary products.

$$\frac{dx(t)}{dt}=Ax(t)+u(t) \qquad (10)$$

For the steady state solution we have:

$$x=A^{-1}u$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ \ldots \end{bmatrix} = \begin{bmatrix} \text{Concentration Enzyme effecting Pelargonidin} \\ \text{Concentration Enzyme effecting Delphinidin} \\ \text{Concentration Enzyme effecting Cyanidin} \\ \text{Concentration Activator of Pelargonidin gene} \\ \text{Concentration Repressor of Pelardonidin gene} \\ \text{Concentration Activator of Delphinidin gene} \\ \ldots\ldots \end{bmatrix} \qquad (11)$$

The measurements can be described as follows:

$$z=Cx=CA^{-1}u$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} \text{Concentration of Pelargonidin} \\ \text{Concentration of Delphinidin} \\ \text{Concentration of Cyanidin} \end{bmatrix} \qquad (12)$$

From this we can obtain the relationships:

$$
\begin{aligned}
&m_1=\kappa_1 z_1\\
&m_2=\kappa_2 z_2\\
&m_3=\kappa_3 z_3\\
&or\\
&m = Kz, and\ z=K^{-1}m
\end{aligned}
\tag{13}
$$

In the above we have assumed the following:

1. The system is at steady state. The expansion to a dynamic system is possible but it is unrealistic for plant colors. It functions in dynamic processes such as blood chemistry and endocrinology.
2. There is an unknown matrix A which we ultimately desire to obtain based upon the measurements made. This is the basis of the system identification problem.
3. There is a constant vector u which is the driver for the system.
4. There is a known relationship between the color elements in the space of colors using anthocyanin elements which can be used to determine the concentration of the anthocyanins. This is the K matrix and we assume that it is a diagonal with known values.

The experimental data approach we use for the system identification process is the microarray. We assume we have a large collection of genes which have been sequenced for the targeted flower. We also assume that we can create a microarray for these known genes and then using the array take samples of many different colors and test them in the array for the gene expressions. We also assume that we can determine expression intensity by measuring the expression intensity in each microarray cell.

In the following Figure 6 we depict a microarray which is composed of cells, one in each row-column pair. Each cell contains a row cDNA sample from a specific phenotype, and a column sample from a specific color, C. The cell then can be measured as to the intensity of the expression of the specific cDNA for each color. These measurements then become the basis of the data set.

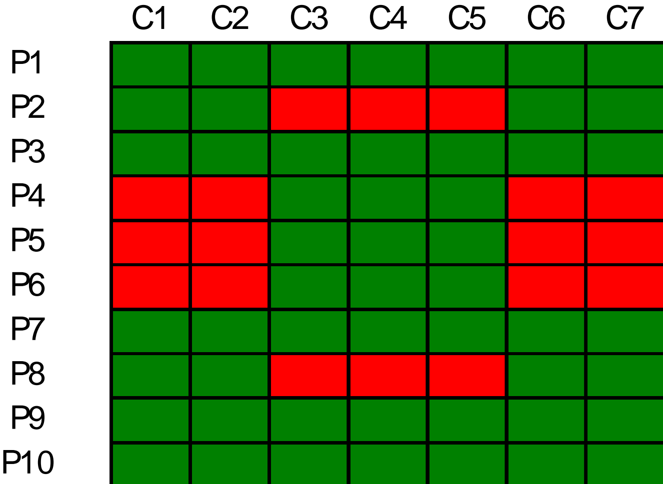


Figure 6: Example of a Microarray output showing activated cells for specific gene and color pairs. The rows are gene elements and the columns are form specific plant flowers.

By using standard means of microarray analysis (see [44], [45], [10], and [39]) as discussed above, the concentration of the protein may be determined. Simply, by having the row represent a know gene from the plant flower, and the columns being for sample from each color, we can measure the relative concentration of the products of genes for each color. This will be a key element in our determination of the system control parameters.

This concentration for the cell is denoted by the variable x. The phenotype color element is processed in a spectrum splitter which uses standard technology to determine the matrix elements of the Red, Green, and Blue elements. Using a color inversion matrix as described previously as G in this embodiment the vector weights for the secondary pathway chemical concentrations are obtained as denotes as vector elements m. Using a concentration matrix inversion processor the concentrations of the separate secondary chemicals denoted by vector elements c, in the case of Figure 7 below.

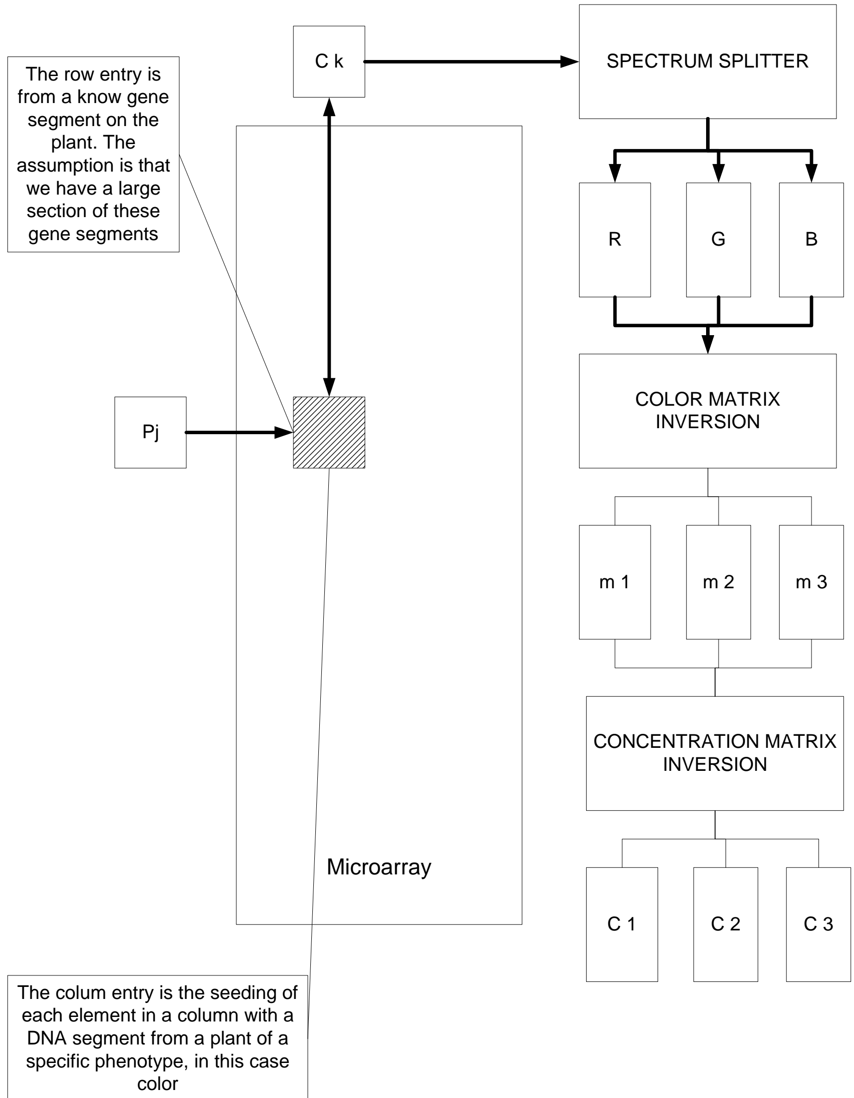


Figure 7: The models for the control of secondary products and their resulting phenotype results and the underlying gene control through enzyme management of the secondary pathway.

Specifically we perform the following steps:

1. Collect samples from various phenotypes, in this case color patches from the flower.
2. Prepare the cDNA from the plant
3. Prepare a microarray using the cDNA versus the various phenotypic color elements.
4. Run microarray analysis to obtain expression for each cDNA and phenotype cell entry.

5. Calculate the gene expression values for each microarray entry and denote the expression as a concentration element x(i,k), which is a measure of the concentration of protein element in row k for each color sample i.
6. Calculate color components for each color entry column and denote these in terms of standard R, G, and B elements.
7. Calculate secondary pathway concentrations of their products, such as anthocyanins, based on the determined R, G, B elements and denote these secondary pathway product concentrations as z(i,k).
8. The above steps provide the basis for the analysis procedure to determine the system control constants, namely the system identification problem.

6.2.7 System Identification

The next step is to use the data obtained to estimate the constants we have assumed exist in the model for the modulator, repressor, activator genes and the control of the secondary pathway.

A set of measurements from the microarray data is collected and are presented in the equations below. The measurements consist of a vector composed of two sets of data; the phenotype elements and the gene expression concentration elements as described above. The equations below depict the measurements consisting of a collection of phenotype elements, m, and gene expression concentrations, x, for the entire collected data set. Namely we have a measurement tuple for every color column denoted by:

$$\{m_1, m_2, m_3, x_1, ..., x_n\} \quad (14)$$

The notation depicting this data set is further described in the equations below. The model employed in this embodiment then can show that the concentrations are related to the unknown expression matrix entries as is shown in the equation below relating x to the elements a of A and the elements of u. The elements a of A are defined in the equations below. Using standard matrix inversion methods the inverse elements are defined in the equations below and this permits the expression of the phenotypic secondary elements in terms of the matrix inverse elements as is shown in the equations (15) below.

$$\begin{aligned} &\{m_1, m_2, m_3, x_1, ..., x_n\} \\ &or \\ &\{m^k{}_1, m^k{}_2, m^3_k, x^k{}_1, ..., x^k{}_n\} \quad k = 1.....N_{measure} \end{aligned} \quad (15)$$

Using the steady state model we obtain for the x values of concentrations the following:

$$x_j = \sum_{i=1}^{N} a^{ji} u_i \quad (16)$$

where we have defined the inverse by the terms shown as follows:

$$A^{-1} = \begin{bmatrix} a^{11} ....a^{1n} \\ .... \\ a^{n1} ....a^{nn} \end{bmatrix} \quad (17)$$

$$\{m_1, m_2, m_3, x_1, ..., x_n\}$$
$$or$$
$$\{m^k{}_1, m^k{}_2, m_k^3, x^k{}_1, ..., x^k{}_n\} \quad k = 1.....N_{measure}$$
$$x_j = \sum_{i=1}^{N} a^{ji} u_i$$

$$A^{-1} = \begin{bmatrix} a^{11} ....a^{1n} \\ .... \\ a^{n1} ....a^{nn} \end{bmatrix}$$

$$m_n(nm) = \sum_{i=1}^{N} \sum_{j=1}^{N} \kappa_{ni} a^{ij} u_j \; : n = 1,3$$

The next step is the calculation of the unknown matrix A from the collected data set of m values and x values specified. The first step in the process is the definition of the unknown elements of the matrix A as a vector of n^2 elements. The method chosen is a least squares fit method using a sequential procedure for obtaining the optimal fit from the data from each of the microcell elements. In the equations below the unknown elements are depicted as an n^2 vector. The objective is to determine an estimate of each entry and the vector estimate is denoted by $\hat{a}$. The least square means is one which minimizes the squared difference between the actual measured data as obtained and what the estimator predicts the data element should be using the most recent best estimate of the values a. This is stated mathematically in the equations below. Using standard mathematical techniques, namely a Newton method for solving the optimality problem, this is shown below as being the equivalent to solving a set of equations of a variable p (see [46] p. 111 for description using the Newton method as is done here).

We thus seek to find an a vector as below:

$$\hat{a}=\begin{bmatrix}\hat{a}_{11}\\ ..\\ \hat{a}_{1n}\\ ..\\ \hat{a}_{n1}\\ ..\\ \hat{a}_{nn}\end{bmatrix}=\begin{bmatrix}a_1\\ .\\ .\\ .\\ .\\ .\\ a_N\end{bmatrix} \quad (18)$$

to minimize the following metric:

$$min\left[\sum_{i=1}^{M}(\hat{m}_i\text{-}m_i)^2+\sum_{j=1}^{N}\left(\hat{x}_j\text{-}x_j\right)^2\right] \quad (19)$$

where we define an h function as:

$$h(a)=\left[\sum_{i=1}^{M}(\hat{m}_i\text{-}m_i)^2+\sum_{j=1}^{N}\left(\hat{x}_j\text{-}x_j\right)^2\right] \quad (20)$$

and we seek a stationary point as follows:

$$\frac{\partial h(a)}{\partial a_n}=p_n(a)=0, n=1...N \quad (21)$$

where the stationary point is defined as:

$$p(a)=\begin{bmatrix}p_1\\ ..\\ p_N\end{bmatrix}=0 \quad (22)$$

Using standard mathematical procedures, the equations in the previous portion of the embodiment may be solved in the following set of equations. The objective is to find the values of a which make the vector p zero. To accomplish this in the following equations a matrix K is obtained by the mathematical procedure depicted in the equations. The estimate of a after k+1 sampled from a microarray depicted in Fig 4 is shown as a function of the estimates after k samples from the same array. This procedure is performed sequentially for all samples in the array as shown in Fig. 4.

We desire to determine the a to solve:

$$p(a) = 0 \qquad (23)$$

We can define a matrix as follows:

$$K(a) = -\left[\frac{\partial p(a)}{\partial a}\right]^{-1}$$

where we define:

$$\left[\frac{\partial p(a)}{\partial a}\right] = \begin{bmatrix} \frac{\partial p_1}{\partial a_1} \cdots\cdots \frac{\partial p_1}{\partial a_n} \\ \cdots\cdots\cdots\cdots\cdots \\ \frac{\partial p_n}{\partial a_1} \cdots\cdots \frac{\partial p_n}{\partial a_n} \end{bmatrix} \qquad (24)$$

with the solution can be sequentially determined as follows:

$$\hat{a}(k+1) = \hat{a}(k) + K(\hat{a}(k))\,p(\hat{a}(k)) \qquad (25)$$

The process for solving the optimization is depicted in the following set of equations. The process commences with an initial guess for the vector a. Then the iterative process begins. A special method of elimination is proposed between steps. The method of elimination is one where genes which are not expressed may be eliminated from the sample. The determination of a non-expression gene can be determined from the process of stating that a gene is non expressing if it does not express or expresses below a set threshold level in any cell. We calculate the K gradient as shown in the equations below. We use the K matrix and calculate the p vector for the next steps as is shown in the equations below. We iteratively use the results of one step to determined the next best estimate. This process proceeds until complete. When complete, the best estimate of the elements of A is determined. The matrix K and the vector p are all evaluated by means of the difference values shown in the following equations.

$$\hat{a}(0)=a^0 \qquad (26)$$

this is the initial guess. The we use the first data tuple which we obtain from the microarray data as it may be normalized:

$$\hat{a}(1)=\hat{a}(0)+A(\hat{a}(0))g((\hat{a}(0)) \quad (27)$$

The difference elements are determined as follows and used:

$$\hat{x}_k(0)-x_{k,measured}(0) \\ \text{and} \\ \hat{m}_k(0)-m_{k,measured}(0) \quad (28)$$

as the data entry element for each of the data elements and where the estimates are calculated using the data collected from the system equations.

6.2.7.1 Modification for on/off A/R Genes

There is a slight modification we must include to deal with genes being on or off. We must return to the beginning to best understand it. Namely if the a's in the system matrix are all constant then by definition the colors remain the same. However, if any one or more of the A/R genes are on or off then we can get variation. We first explore the implications of this and then we modify the estimator process accordingly. Let us review our model assumptions:

1. We assume that the genes directing the secondary pathway are always functioning.

2. We assume that the constants in the gene expression model and secondary control are all constant and remain so.

3. We assume that the A/R gene may be on or off. They are controlled via some tertiary process yet to be determined.

Thus, we can consider the example of a three gene system with two A/R gene per expression gene we have 4X4X4 possible states. This means we have 64 possible color states. If we have n A/R genes per expression gene and we have m expression genes we have 2^{nm} possible color states.

Now the above algorithm is a least squares estimate algorithm given an A/R gene state. We now propose a model where we first estimate the state of the A/R genes and then given that state we use the least squares approach to estimate the a values which remain. Thus if a specific A/R gene is in a 0 state we then zero out its effecting a value and estimate the remaining a values as we would have done before. The mathematical analysis to justify the algorithm uses the MAP (maximum a posteriori) estimate approach (see [50]). Specifically, we maximize the following:

$$\frac{\partial \ln\ p_{a/z}(A/Z)}{\partial A} = 0 \qquad (29)$$

[50] has shown the equivalence to the minimum mean square estimator (MMSE) approach or the Bayesian analysis. Thus we write the MAP estimator as:

$$\frac{\partial \ln\ p_{a/z_C,z_D}(A/Z_{Continuous},Z_{Discrete})}{\partial A} = 0$$

$$is;$$

$$\frac{\partial \ln\ p_{a/z_c}p_{z_c/z_D}}{\partial A} \qquad (30)$$

clearly from the above we can separate the two optimizations. Namely this tells us that we can, first estimate the binary values of the on/off states of the A or R genes. Then we can use the standard approach to obtain the continuous, C, elements, in our current case the a values. The best estimator for the binary part is a standard MAP estimator using a threshold. We can perform this task by examining each microcell entry for it being active or inactive.

The algorithm for the calculation is shown in Figure 8. We perform the functions as we stated above.

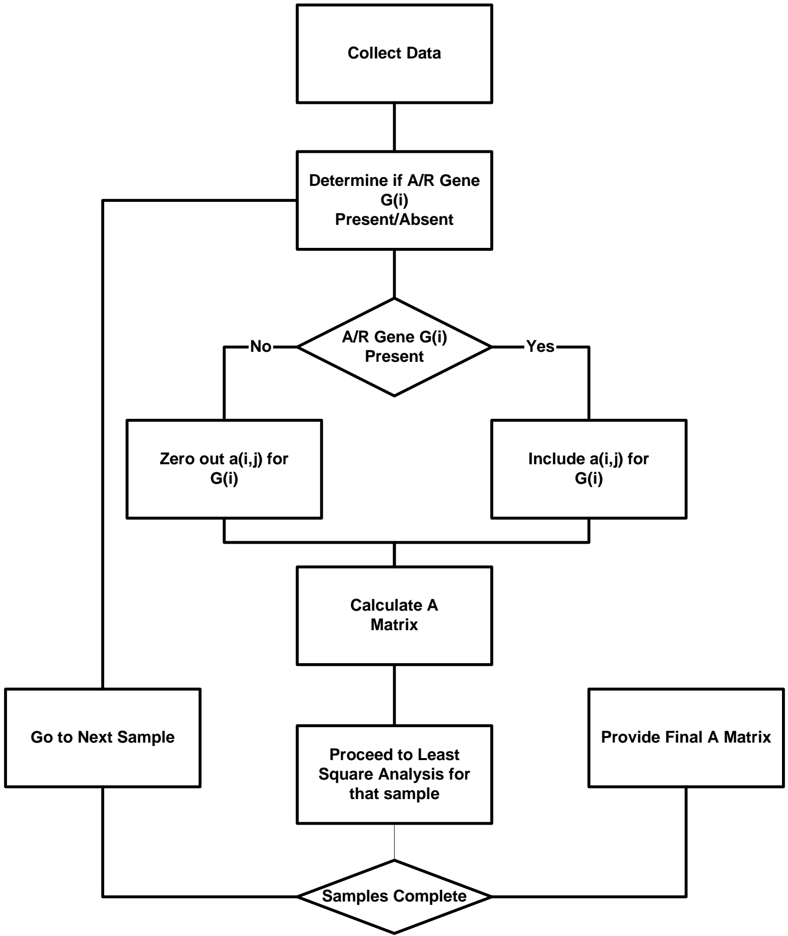


Figure 8: Algorithm for the calculation of the testing and elimination of "zeroed" gene effects.

6.2.8 Estimation versus Identification

The method of estimating the structural elements of the gene expression can be structured using a standard set of methodologies. The approach was applied to estimating the constituent chemical concentrations of the upper atmosphere, namely the inversion problem, using transmitted light as the probe mechanism. In this case we seek to estimate the gene expression matrix using the concentrations of secondary chemicals as expressed in color concentrations. This is in many ways a similar problem.

6.2.8.1 The Model

Let us consider a six gene model, two color modifying genes and four control genes, two each. The model is as follows. First is a general linear model for the gene production:

$$\frac{dx(t)}{dt} = Ax(t) + u(t) + n(t) \quad (31)$$

Then the entries are as follows:

$$A = \begin{bmatrix} a_{11}..a_{12}..a_{13}..0..0..0 \\ 0...a_{22}..0...0...0...0 \\ 0...0...a_{33}..0...0...0 \\ 0...0...0...a_{44}..a_{45}..a_{46} \\ 0...0...0...0...a_{55}..0 \\ 0...0...0...0...0...a_{66} \end{bmatrix}$$

$$and$$

$$u(t) = \begin{bmatrix} u_1 \\ ... \\ u_6 \end{bmatrix} \quad (32)$$

and we assume a system noise which is white with the following characteristic:

$$E\left[n(t)\right] = 0$$
$$and \quad (33)$$
$$E\left[n(t)n(s)\right] = N_0 I \delta(t-s)$$

Now we can define:

$$A = \begin{bmatrix} A_1...0 \\ 0...A_2 \end{bmatrix} \quad (34)$$

Where we have partitioned the matrix into four submatrices. This shows that each gene and its controller are separate. Now we can determine the concentrations of each protein in steady state as follows, neglecting the Gaussian noise element for the time being:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = -A_1^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

$$and$$

$$\begin{bmatrix} x_4 \\ x_5 \\ x_6 \end{bmatrix} = -A_2^{-1} \begin{bmatrix} u_4 \\ u_5 \\ u_6 \end{bmatrix} \qquad (35)$$

We argue that finding either the matrix elements or their inverse relatives is identical. Thus we focus on the inverse elements. Now the concentrations of the anthocyanins are given by the 2 x 2 vector as follows:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} c_{11}...0...0...0...0...0 \\ 0...0...0...c_{24}...0...0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = Cx \qquad (36)$$

The color model remains the same.

6.2.8.2 The Estimator Model

The system model is as follows. Let us begin with a model for the vector a that we seek:

$$\frac{da(t)}{dt} = 0 : where$$

$$a(t) = \begin{bmatrix} a_1 \\ ... \\ a_5 \end{bmatrix} \qquad (37)$$

In this case we have assumed a is a 5 x 1 vector but it can be any vector. The measurement system equation is given by:

$$z(t) = g(a,t) + w(t) \quad (38)$$

Where z is an m x 1 vector. In this case however we have for the measurement the following:

$$z(t) = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ x_1 \\ \cdots \\ x_6 \end{bmatrix} = g(a,t) + w(t) \quad (39)$$

We now expand in a Taylor series the above g function:

$$g(a,t) = g(a_0,t) + C(a_0,t)\left[a(t) - a_0(t)\right] + \frac{1}{2}\sum_{i=1}^{N} \gamma_i \left[a - a_0\right]^T F_i \left[a - a_0\right] + \ldots \quad (40)$$

Where we have:

$$C = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_n} \\ .. & .. & .. \\ \frac{\partial g_m}{\partial a_1} & \cdots & \frac{\partial g_m}{\partial a_n} \end{bmatrix} \quad (41)$$

Thus we have for the measurement:

$$z(t) = C(t)a(t) + \left[g(a_0) - C(a_0)a_0(t)\right] \quad (42)$$

We now use standard Kalman theory to determine the mean square estimate;

$$\frac{d\hat{a}(t)}{dt} = P(t)C^T(t)K^{-1}(z - C(t)\hat{a}(t))$$

$$where$$

$$\frac{dP(t)}{dt} = -P(t)C^T(t)K^{-1}C(t)P(t) +$$

$$\sum_{i=1}^{N} PF_i P\gamma_i^T K^{-1}(z - g(a_0)) \quad (43)$$

$$where$$

$$K\gamma(t-s) = E\left[w(t)w^T(s)\right]$$

In discrete time we have the equation:

$$\hat{a}(k+1) = \hat{a}(k) + PCK^{-1}\left[z(k) - \hat{z}(k)\right] \quad (44)$$

This is identical to the equation we derived from the Newton method.

6.2.8.3 Model Variants

We now want to consider several variants on how color can be generated. In our model we assumed that Expression constants and drivers remained the same throughout but that they were turned on and off thus generating differing colors. However there are many possibilities,

1. Expression Constants Vary: In this case the a values vary from color to color. From a gene expression perspective we cannot find this an attractive alternative. However it may be conceivable that there are secondary controller A/R genes which may be playing a role which may be unidentified at the time of the experimental analysis.

2. Expression Drivers Vary: The u values we have used to be the steady state drivers may be affected by various factors, including local factors such as plant acidity and location. We have assumed we know these variables. This assumption is based upon some past experimental analyses. However these may vary and must be taken into account. To do so we can expand the model to estimate them as well.

3. Expression Constants are on/off: This is assumed in the model we have developed herein. This model assumes that all variables are constant and that we have just an on and off process of A/R genes.

4. Expression Drivers are on/off: This is an intriguing alternative with no known physical embodiment but it may be the case from time to time. However the effect is the same for A/R expression as if we assumed the constants went on or off.

Any combination of the above may also occur.

6.2.9 Conclusion

In this section we have presented an interesting approach to study with respect to gene expression and ultimately the control of gene expression. The phenotypes are quite obvious in flower colors and in addition the hybridizing which has led to a wealth of examples has been done just in the past one hundred years. Also we have a reasonably clear understanding of the underlying species and we can readily assess the complexity of the species DNA structure.

There are three problems for which this approach applies:

1. **Analysis:** In the analysis problem we assume we know the expression dynamics and the secondary production model. Given those two we determine what color we get. We develop this in detail.

2. **Identification:** The identification problem is one in which we know the secondary processes, we have many color samples and we know the protein concentrations which yield each color. Then we ask how do we determine the A matrix for the gene expression?

3. **Design:** This problem is of significant interest. We seek a desired output state, color in our example. We know the gene expression dynamics and the secondary model. We then show how to modify the gene expression model to obtain the desired output.

We also have a well defined and understood set of pathways that give rise to the phenotype. We further know the effecting proteins and enzymes. We also know the gene which affects the proteins in question. Finally we have well accepted models for the expression of the genes and we can use generally accepted models for the dynamics of gene expression.

This has led us in our final section to a modeling of gene expression as a set of definable dynamic systems. We have used a certain set of those systems to discuss examples. However certain key questions remain:

First, what are the dynamic models which can adequately and correctly describe the abrupt coloration of the flowers? We have a good understanding of many of the

unstable dynamic systems models which can describe such phenomenon but what is the relationship between what occurs in the cell and what the models describe?

Second, we have used an ensemble approach versus the microbiologists' time approach to modeling the system. We have posited an equivalency based upon the Ergodic Theorem, which states the time average and ensemble average are equal. However there is no experimental proof of this fact.

Third, in any systems approach, we always look at observability and controllability. Observability concerns whether we can see the outputs knowing the system model and can we predict the initial condition. This must be validated experimentally. Controllability is simply can we drive the system to a desired state with a control function. The controllability question goes to the heart of flower color design. If we accept the validity of our models the answer appears to be determinable for any set of defined pathways.

Fourth, we have suggested a microarray approach to estimating the coefficients of the dynamic system. This is one of many possible techniques. The first part we should do is address this from an experimental perspective. Namely perform the microarray analysis. The second part is to investigate alternative methods of solving the system identification problem via alternative bench based validation tools.

Fifth, specific phenotypic design must be considered in more detail and experimentally validated.

Sixth, we use a stochastic model for the expression and pathway analysis. We used this as a way to account for dimensions we could not include because they were expressed at too low a level or because we had no knowledge of their existence. Thus we argued that noise may be true random processes or the aggregation of currently unknown tertiary processes. Experimental validation of this modeling element must be performed.

Seventh, can this approach be carried over to any other cell line? The answer we believe is yes it can and readily. What we have done herein is to focus on phenotypic characteristics and ones which are readily characterizable by well understood pathways. Such systems exist in many other biological systems including the human.

7 CLASSIC GENETICS AND HYBRIDIZING OF HEMEROCALLIS

The hybridizing of the Genus Hemerocallis has been performed for just over a hundred years and within the past thirsty there has been an explosive growth in the effort. The hybrids being introduced have a significant amount of variation in color and form and they flower is changing in ways that have been generally unpredicted a generation ago. There are various color and more importantly various shapes. In this paper we address the hybridizing from a classic Mendellian viewpoint and then look at many of the hybridizers over the past hundred years. Our goal is to provide some insight into the hybridizing process by looking at the fundamental scientific basis and then looking at the approach actually by many of the well respected hybridizers.

7.1 INTRODUCTION

The process of hybridizing plants is an old one and it has been developed over centuries if not over millennium. The most recent first step in improving the process with some scientific basis was the use of Mendellian methods. This paper will review the classic Mendellian approach, the classic breeding approaches to seek out traits and then we will review some methods employed in current day hybridizing.

Hybridizing as is performed in the genus Hemerocallis is in many ways akin to the classic approaches use in farm crops as well as horse breeding. The goal of Hemerocallis hybridizing depends upon the hybridizer and most current hybridizers are interested in incrementally adding to the stock of "different looking" flowers. Very few of the current hybridizers are approaching the genus with scientific inquiry but the data is available with the AHS database to seek out some of these characteristics.

The approach we take herein is focused on looking at the process as a Mendellian. There is in this paper no focus on modern genetics but as we have discussed early on the Mendellian approach has its limitations. What may appear as a surprise for a Mendellian may be just a natural progression for a current generation plant geneticist.

Hybridizing is also an area where technique and technology can blend. For many hybridizers their approach is more of an art than a science. Underlying the basis of hybridizing is the genetic makeup of the plant but the genetics can be so complex and possible not understandable to many of the hybridizers. Their approach is to build upon the work of others. Thus hybridizing as currently practiced is an art of good guessing and good crosses.

In this Paper we further develop some of the classic Mendellian methods of analysis and synthesis. As we have stated elsewhere these are somewhat crude methods which may apply to certain gross characteristics as pea color and size but when applied to specific

flower color and complexity have been found wanting. The methods may still have significant use and utility when trying to induce more extensive branching as an example, whereas the ability to control such characteristics as eyezone the techniques may get called into question.

The objectives of this paper are:

1. Summarize the insights of the Mendellian approach and to understand its limitations, to better see what the principles are and to see where the hybridizers have been focusing in applying these methods, if at all.

2. To understand some of the implications of Mendellian genetics to see how inbreeding may result in the introduction of certain traits into a plant line lacking such traits. These methods have been successful in the many types of plants used for food such as grains.

3. To establish a process and methodology for breeder or hybridizers. This means the setting of reasonable goals and the methods which may be employed to achieve those goals.

4. To review and understand the various standard methodologies used I breeding and their advantages and disadvantages.

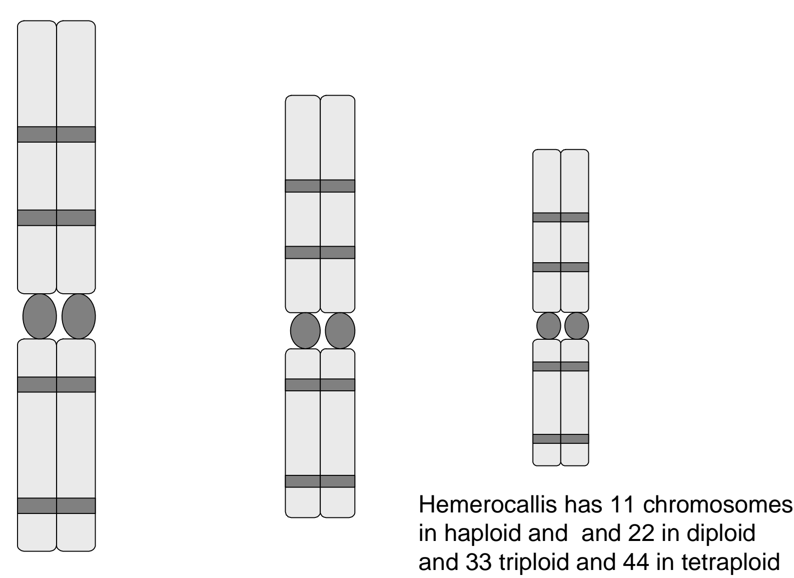
5. To look at the evolution of hybridizing in the United States of Hemerocallis and to attempt to understand how the principle developed herein were and will be applied.

7.2 Classic Mendellian Concepts

The Classic Mendellian approaches are all based upon the Mendel model of the gene and traits. Unlike what we look at in modern day genetic analysis, the classic Mendel or Mendellian approach looks at chromosomes and then at genes. The gene then becomes the effector of the characteristic we are trying to duplicate, enhance, eliminate, or whatever. The Mendellian paradigm of the "gene" is a gross concept that links a phenotypic characteristic such as color, height, branching, or whatever, to a specific Mendellian gene. There is assumed to be a one to one and un-modifiable relationship between this gene and the phenotypic character.

In Mendellian analysis we assume that there exists a gene on a chromosome which provides some characteristic, say yellow petals. Hemerocallis has 11 chromosomes with a diploid being the normal configuration. There is no sex chromosome as in humans. There are triploids with 33 chromosomes and tetraploids with 44 chromosomes. But the species has 22 in each cell. We characterize this as shown below.

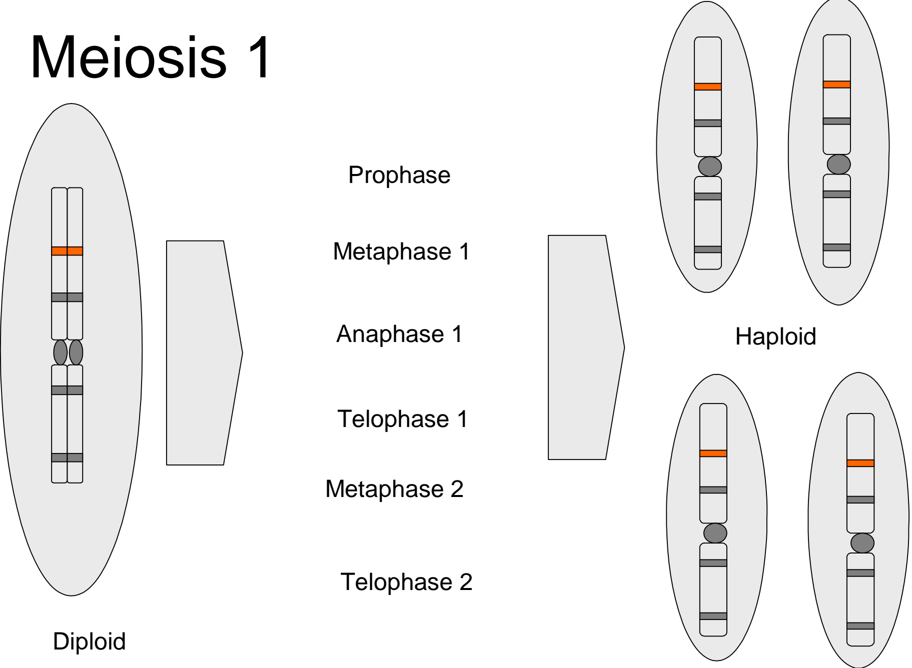
TABLE 1 CHROMOSOMES AND GENES



Now each chromosome with some gene segment has some controlling characteristic, such as a gene for the color orange. In the process of meiosis in the sex sells the chromosome pairs split, some even cross over, and a mixing and matching of chromosomes and genes are made. Our intent is not to provide a detailed summary of Mendellian analysis but to focus on the key points which will be used to continue our analysis.

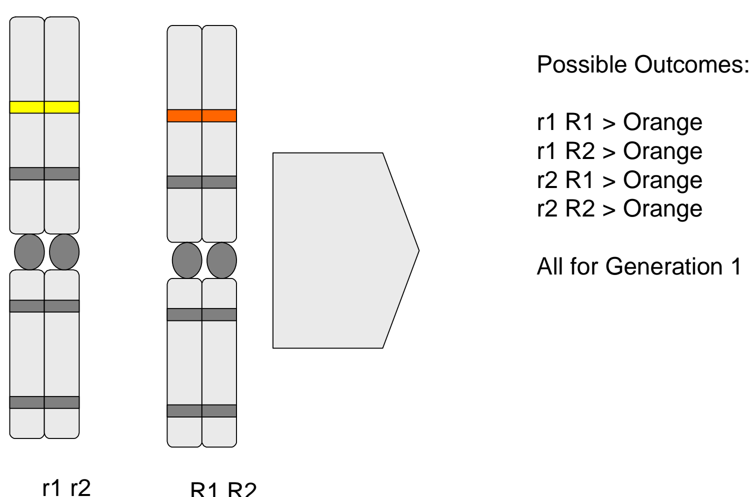
When a plant creates a pollen grain or a female oocyte they are products or meiosis. And these cells are haploid, namely only one copy of the chromosome. Thus is we have two orange genes, one on each chromosome, and then in meiosis we end up with a male and female haploid cell each having one orange gene.

TABLE 2 MEIOSIS STEP 1



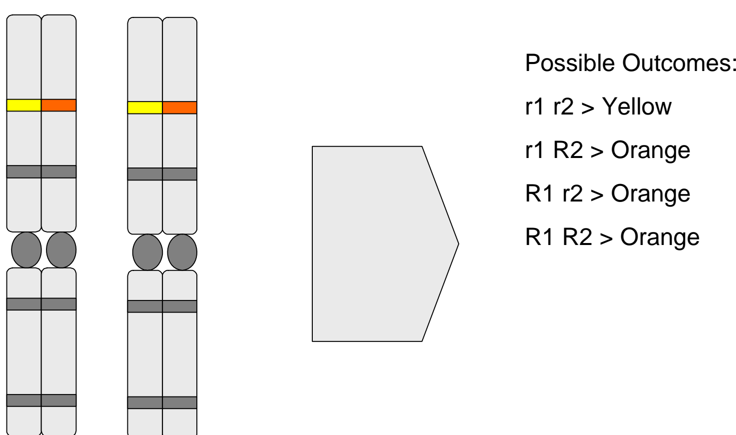
If, however, we take a yellow plant, allow it to create the haploid cells via meiosis and take a pure orange plant, let it create haploid cells via meiosis and then cross these plants we get what we see below.

Namely there are four possibilities, each equally likely; we have a yellow with an orange in any one of the ways shown below. Thus in what is called the F1 generation we only get orange plants since the orange gene is dominant and each of the F1 plants have the same genetic makeup, a haploid with a yellow and a haploid with an orange.



Now we go to the F2 generation. This is the offspring of the F1. Remember that all F1 have same gene structure, a yellow and an orange gene. These break apart in meiosis and combine again when the plants are fertilized. The net result in the off spring in F1 is a set of chromosomes with a yellow and orange chromosome. When they split there is a possibility of the off spring of the off spring in the F2 to have two yellows which means yellow or one of each yielding orange or a pure orange. Thus with one gene we find that a dominant gene will give $1/4^{th}$ with the recessive and $3/4^{th}$ with the dominant color. We show this below.

TABLE 4 MEIOSIS STEP 3



7.3 BASIC MENDELLIAN GENETIC ANALYSIS

Before proceeding to the issues of hybridizing, we will consider some basic Mendellian genetics as a framework for helping to understand how to perform the hybridizing tasks. Let us make the following assumptions, which are what Mendel made in his experiments and analyses:

1. There exists a construct on the chromosome called a gene.

2. Let us assume that chromosomes come in pairs and that each chromosome has a gene which has the effect that we are trying to analyze. This in a Hemerocallis species there are 11 chromosomes and they come in pairs so there are 22 chromosomes and there are genes on each one of the chromosomes.

3. The gene can be one of several types; we generally assume that there are just two types of genes. We label these say A and a and two genes.

4. A gene yields a phenotypic characteristic which we can observe. Gene A yields one type of phenotypic characteristic and gene a another. We assume that these characteristics are clearly distinct. They may be the presence or absence of an eyezone in a flower.

5. The gene controls a characteristic of a flower or plant and that the gene is the sole control element of that characteristic.

6. That there exist dominant and recessive genes. The dominant gene if present yields the phenotype consistent with that gene whether there is one or two of those genes present.

We must understand, however, that the Mendellian gene construct differs widely from the current understanding of a gene in many ways. We will return to this in later papers.

To understand the world view of Mendel, one must understand that he worked with peas primarily and he did extensive crossing and observed clear and delimited traits. There were limited colors and limited shapes. It would be akin to a daylily leaf, being grass-like or broad, long and heavy, H minor versus H fulva. It was clear what the difference was.

The Mendellian analysis did not try to account for subtle and sophisticated variations in form, shape, color of the highly hybridized daylily.

7.3.1 Simple Crosses

We begin with a simple cross between two plants. Let us assume that there is a characteristic which can take one of two states. We further posit that the character is controlled by a single "gene" and we call that gene A or a, depending on the state that is taken. Now let us assume that we have two plants; plant 1 and plant 2. Furthermore we somehow "know" that Plant 1 is AA and Plant 2 is aa, namely Plant 1 has two genes on the two chromosomes that are both type A and likewise for plant 2 they are both a. Then we ask, what happens if we were to cross breed these two plants. Let us assume for example that A yields no eyezone and a yields an eyezone.

Before proceeding we must say a bit more about the gene mechanism. We say that the gene A is a dominant gene and that the gene a is a recessive gene. What do we mean by that? We mean that if one or both of the chromosomes have an A gene then the characteristic generated by A will be in evidence. If, however, the plant were to have two a genes then the character related to a would be in evidence. Dominant means that as long as there is at least one then its effect is evident. Recessive means that no matter what we can only have the a gene present.

Understanding the current world of transcription, we know now that on a gene paid on two bound chromosomes of DNA, the reading of the gene to the RNA is done on only one of the genes, never both. Thus this currently understood fact may help explain what happens. If A is on one or both of the chromosomes, then A forces the transcription process, no matter what, leaving an a gene un-transcribed.

Let us go back to this simple cross. We take the two genes of the recessive eyezone and place them across the top. We take the two genes of the pure dominant and place then along the side. Then the possible outcomes when we combine these two through breeding or hybridizing are shown in the Table below.

	a	a
A	aA No Eye	aA No Eye
A	aA No Eye	aA No Eye

Before looking into the details let us analyze this methodology. Each of the two parents has two chromosomes and on each chromosome there is a gene. In the dominant parent this means that we have a gene A on one chromosome and a gene A on the other. These are identical genes but NOT the same gene. In the Mendellian analysis it

assumes that either gene may act. In a similar manner we have the same situation for the recessive gene, a, and there are two of them. Thus when the parents combine their chromosome into a new plant, the new plant has one chromosome from each parent. This simply means it gets an A from the dominant and an a from the recessive. In the above the row across the top lists all possible genes from the recessive and the column to the left all possible chromosomes from the dominant. Even if they are both identical they represent two genes, one from chromosome 1 and one from chromosome 2. Another way to look at this is to write the Table as below:

	a1	a2
A1	a1A1 No Eye	a2A1 No Eye
A2	a1A2 No Eye	a2A2 No Eye

Which is identical to the above except now it shows gene and chromosome.

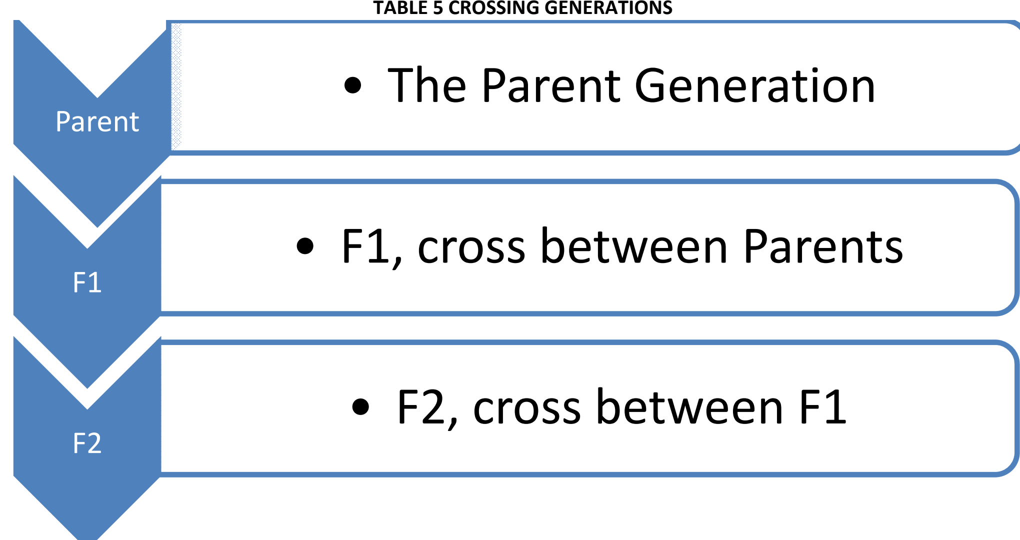
We can now take this one step further as an attempt to clarify the detail. If we call the parent in the top row as 1 and the parent in the left column as 2 we have gene:chromosome:parent as a tuple. This we show below:

	a:1:1	a:2:1
A:1:2	[a:1:1, A:1:2] No Eye	[a:2:1 , A:1:2] No Eye
A:2:2	[a:1:1 , A:2:2] No Eye	[a:2:1 , A:2:2] No Eye

The details here are complete. Each heading, row or column, species a parent gene and its sours, namely what chromosome and what parent. It also specifies what specific gene it is. This level of detail will greatly assist in complex analyses.

Now the result of this crossing yields what we call the F1 generation, the off -spring from two pure bred plants. Let us define the generations since we will use them again.

TABLE 5 CROSSING GENERATIONS



Now we would know they are pure bred after the fact if and only if there were no eyezones, in any offspring. That of course is not practical. We can statistically say that one parent is AA if a large number are without eyezones. We will come back to that later.

Now assume we take one of the F1 offspring and breed them with another F1 offspring. What do we get? Well, all of them have aA gene pairs on their chromosomes. Thus if we create the same map as before this time we have:

	a1	A2
a1	a1a1 Eye	a1A2 No Eye
A2	a1A2 No Eye	A2A2 No Eye

Thus in this simple case we have a chance of one in four, 1:4, or 25% that there will be an eye zone.

Now what does this tell us about hybridizing daylilies. Frankly, there is very little. Mendel had peas, and he was looking at peas all one color, one gene one phenotype. There was no mixing, no complicated gene control. There could be a simple control of a gene and a phenotypic characteristic.

For example, if we had a daylily with an eyezone and bicolor and no eyezone was dominant as was an non bicolor, then the table below predicts the result. This means

that we have two genes, one pair being B and b and the other A and a. The b gene is for a bicolor and the a gene for an eyezone. This is the classic Mendel analysis. We show the Table below for the example of an A and B gene with a dominant A and dominant B and the recessives a and b.

To perform this analysis let us assume we start with two purebreds as before, but this time we have:

Plant 1: The following genes are available; A, A, B, B (no eye zone and no bicolor)

Plant 2: a, a, b, b (eyezone and bicolor)

The results for the F1 generation are as follows:

	AB	AB	AB	AB
ab	AaBb	AaBb	AaBb	AaBb
ab	AaBb	AaBb	AaBb	AaBb
ab	AaBb	AaBb	AaBb	AaBb
ab	AaBb	AaBb	AaBb	AaBb

Note we could clarify this by noting that the genes across the top are A1, A2, B1, and B2 all from the dominant parent. Likewise for the column we would expect to see the same.

However, this is not the case. Go back and look at the species and then look at the hybrids. How does one go from here to there? That is a key question. Genes are being expressed differentially in various was and the control of those expressions varies across the sepal and petal. That is an issue we wish to explore.

Now cross the parent with any F1. There are two possibilities; the dominant parent or the recessive parent. First the dominant crossed with F1.

	AB	AB	AB	AB
ab	AaBb	AaBb	AaBb	AaBb
aB	AaBB	AaBB	AaBB	AaBB
Ab	AABb	AABb	AABb	AABb
AB	AABB	AABB	AABB	AABB

The row across the top is as was before. Now, however the column on the left represents the possible 2-tuples from the F1 crosses. Note that we can combine the 4-tuple four different ways two at a time. Again as with F1 all are controlled by the

dominant genes. We would observe no difference. Now cross F1 with the fully recessive. We obtain:

	ab	ab	ab	ab
ab	aabb	aabb	aabb	aabb
aB	aabB	aabB	aabB	aabB
Ab	AAbb	aAbb	aAbb	aAbb
AB	aAbB	aAbB	aAbB	aAbB

This yields:

1. 4 with eyezone and bi-color
2. 4 with eyezone
3. 4 with bi-color and
4. 4 as the dominant

This is called a backcross.

7.3.2 Complex Crosses

One of the more complex issues arises when we consider tetraploids. In Hemerocallis, with the induction of a tetraploid, each parent has four chromosomes and thus four genes. The gamete cells have in them two chromosomes instead of the one so that when they combine the resulting cell again has four. This adds a bit of complexity. Now we can have the following if we have two homozygous cells:

Pollen Cell (Plant 1): A:1:1, A:2:1, A:3:1, A:4:1

Ovary Cell (Plant 2): a:1:2, a:2:2, a:3:2, a:4:2

and they can be combined two at a time. Thus we can see:

(A:1:1, A:2:1), (A:1:1, A:3:1), (A:1:1, A:4:1), (A:2:1, A:3:1), (A:2:1:, A:4:1), (A:3:1, A:4:1)

and the same for the recessive plant;

(a:1:1, a:2:1), (a:1:1, a:3:1), (a:1:1, a:4:1), (a:2:1, a:3:1), (a:2:1:, a:4:1), (a:3:1, a:4:1)

Using this methodology we can be certain that we track all possible chromosome pairings. This becomes dramatically more complex by just adding another trait.

7.4 Genes, Dominance, Color

Having gone through several examples in the above crosses we may ask if there are genes for dominance of certain colors, shapes, and variegations in Hemerocallis, specifically we would start with color. In a paper by Joanne Norton[41] in the Daylily Journal in the 1970s, the author makes a set of statements, regrettably with absolutely no scientific basis in fact, concerning the Mendellian genetics of Hemerocallis and hybrids. It is regrettable that such is done because she may very well have had some basis for her statements other than purely anecdotal and that would have helped greatly. However Norton appears to be somewhat knowledgeable but in her rather heavy handed statements, without any evidence presented, calls all her work into question[42].

Notwithstanding we try to summarize her results and to comment based upon our experience. The reason for this attempt is the otherwise total lack of any discussion regarding the hybridized version of the genus. Recently Hart has re-presented the Norton work in a more readable and up dated format which is helpful. However, Hart just represents the Norton work and does not seem to have added any fundamental experimental data analysis. However, we do believe that it is worth the exercise to study Norton because she presents questions in a Mendellian manner which can ultimately be proven correct or not. Yet we also have shown that the Mendellian approach to color and pattern formation is greatly wanting. It totally fails to address the epigenetic issues and also fails to deal with the secondary pathway problem.

To ascertain the true relationships, however, one must perform a detailed experimental study to ascertain the true relationships and dominance. In addition, as we had discussed herein, the color question is quite complex since it is gene expression through secondary pathways and this complex set of relationships transcends the simplistic single gene theory espoused by Norton.

(1) Color

Assertion 1: There is a dominant gene for pink, P, and the recessive gene p is homozygous in all yellows, namely pp.

[41] Norton received bachelor's, master's and doctorate degrees in botany from The Ohio State University. Following her graduate work, she was on the faculty at the University of Texas for about two years. http://www.wheresoursquirrel.com/cgi-bin/fish/YaBB.cgi?board=live;action=display;num=1121912943

[42] See Norton p. 2 where she states "my records would be much more useful if I had kept descriptions of all the seedlings…" The fact that we are making conclusions on a selected set invalidates any and all claims.

Assertion 2: There is a dominant gene for yellow, Y, which is in all yellow plants.

Assertion 3: Y and P may or may not be "alleles", namely on the same chromosome.

Assertion 4: There is a dominant gene for red, R.

Assertion 5: All cream, pale yellow, medium yellow, gold and orange plants have Y but no P. Yellows are YYpp or Yypp.

(2) Form

Assertion 1: There are six forms or patterns of flowers[43]:

- Solid
- Eyezone
- Dusted
- Bicolor
- Bitone
- Edged

Assertion 2: For the pattern to be expressed there must exist a gene for that pattern and it must dominate.

Assertion 3: All the patterns are expressed if and only if the P gene is present.

Assertion 4: The color of the pattern is controlled by the same modifiers of P that affect the color of a solid color containing P.

Assertion 5: More than one pattern can appear on a flower. Although two patterns may appear many have only one visible.

Her discussions on patterns are totally baseless. It is know from the early work of Turing and others that patterns are highly complex genetic mechanisms, somewhat akin to fractals. They are highly interlinked epigenetic mechanisms which create the pattern and color variations in what may appear to be an almost random form but have true structure. We will defer this discussion to a latter paper.

[43] There is no basis other than observation for this assertion. The paper by Turing addresses the issue of genetic patterning. Turing may have provided a detailed underlying methodology to prove her assertion or to disprove it.

Norton continues with dozens of anecdotally based assertions in the preceding manner. Hart has done a superb job in summarizing these and we will use the result of Hart rather than belaboring the Norton approach[44].

The summary by Hart of Norton is as follows:

TABLE 6 COLOR AND DOMINANT AND RECESSIVE

Gene	*Dominant*	*Dominant Effect*	*Recessive*	*Recessive Effect*
Yellow	Y	Yellow color	y	Mellon color
Pink	P	Pink or Lavender	p	not pink or lavender
Red	R	red	r	no red
Pink Influencing	IP			
Lavender Influencing	IL			
Drabbiness	D		d	
Muddiness	M			

From the above Table we can present a genetic profile for flower color as follows:

$$\left[\left(G^1_{Yellow}G^2_{Yellow}\right),\left(G^1_{Pink}G^2_{Pink}\right),\left(G^1_{\text{Red}}G^2_{\text{Red}}\right),\left(G^1_{PI}G^2_{PI}\right),\left(G^1_{LI}G^2_{LI}\right),\left(G^1_{Drabby}G^2_{Drabby}\right),\left(G^1_{Muddy}G^2_{Muddy}\right)\right]$$

$$where$$

$$G_{Yellow}=\begin{cases}\text{Y, dominant}\\ \text{or}\\ \text{y, recessive}\end{cases}$$

The above does not imply any chromosomal relatedness or linkage. In addition there is no statistical basis for any of the above it is solely anecdotal. Furthermore there is no genetic or secondary pathway for any of the above. In fact the Norton Conjecture is just

[44] See Hart http://www.hartsdaylilies.com/index.htm

that, anecdotal conjectures which in light of their being anything else remain. In some way they remain a paradigm to be proved or disproved.

The following Tables are modified from Hart[45]. They are allegedly based upon Norton as well. The first is color:

TABLE 7 COLOR AND GENES (NORTON MODEL)

Color	*Gene Profile*	*Dominant Gene*	*Recessive Gene*	*Secondary Pathway Element*
Melon	{(y,y),(p,p),(r,r)……}	None	y p r	lycopene and no anthocyanins
Yellow	{(Y,X),(p,p)(r,r)…..}	Y	p r	beta carotenes no anthocyanin
Clear Pink	{(y,y),(P,X),(r,r),(IP,X),(d,d)…}	P IP Hart also posits it may be Y,X as well as yy	y r d	lycopene and delphinidin
Muddy Pink	{(yy),(P,X),(r,r),(IP,X),(D,X)…}	P IP D	y r	NA
Peach, Apricot, Copper	{(Y,X),(P,X),(r,r),(IP,X),(d,d)…}	Y P IP	r d	NA
Duff, Tan, Brown	{(Y,X),(P,X),(r,r),(IP,X),(D,X)…}	Y P IP D	r	NA

Hart also introduces two more genes which he argues control secondary pathways via gene enzymatic regulation. These are summarized below.

[45] See Hart and also Eder PhD Thesis Munich, http://deposit.ddb.de/cgi-bin/dokserv?idn=963026275&dok_var=d1&dok_ext=pdf&filename=963026275.pdf The Thesis is in German but with a modicum of German and a good base in chemistry it is approachable.

TABLE 8 ENZYMES (PROTEINS) AND GENES

Enzyme (Gene Product)	Color	Dominant	Recessive
F3'H	Red (cyanidin)	R	r
F3'5'H	purple (delphinidin)	P	p
FHT	flavones	E	e
FLS	flavones	L	l

We have discussed these pathways in detail elsewhere. There are issues regarding rates of enzyme production and the like which may dramatically modulate these pathways. Hart does not discuss these at all.

Hart then proceeds to layout colors and these additional genes. We assume that we would have to expand the genes to account for the two controlling the secondary enzymes proposed by Hart. The Colors and the Putative Norton Genes[46] as well as Hart genes are shown below[47].

[46] See Hart, Genetics of Daylilies, http://www.hartsdaylilies.com/genetics.htm

[47] Hart uses the term "No" and it is not at all clear what he means by that. There are other entries which are blank and then there are ones which have a pure negative term. One is left wondering from the Hart presentation but one need only look at the chemistry to clarify. We do that elsewhere.

TABLE 9 ANTHOCYANIN AND GENES

Delphinidin	*Cyanidin*	*Quercitin*	*Dominant*	*Recessive*
X		No	P E L	r possible l
X		X	P E R L	
X	No		P E	r
No	X		E R	p
X	X		P E R	

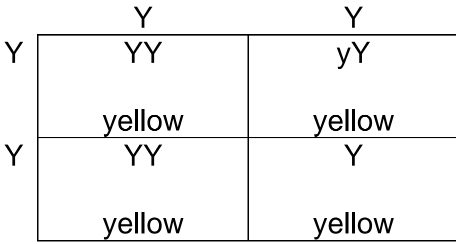
Now it is possible to prove or disprove the above conjectures. All one needs to do is perform the crosses and perform a detailed statistical analysis.

Before proceeding we will use the Norton-Hart model to discuss what could and possibly should have been done to validate the assertions. Let us assume we can take two flowers, a Yellow and a Melon. We know from the Norton Assertions that we have (y,y) for melon and (Y,X) for Yellow. All the other genes are recessive and identical and thus we should have a simple analytical case if we breed them.

First we should self cross the Yellow. This will tell us if we have (Y,Y) or (Y,y). Remember we have the following two cases.

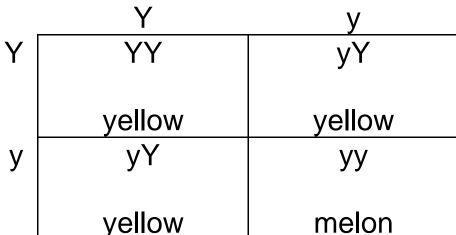
7.4.1 Case 1 YY Parent Self Crossed

By self crossing the parent we should have all yellow offspring. This is shown below



7.4.2 Case 2 yY Parent Self Crossed

By self crossing we should have 25% melon. The 25% melon gives us the desired result.



We can plot the probability of these two events as below. Namely if we have a YY and we self cross it there should be no melon flowers at all. If we have a yY and we self cross it then there fraction of melon is 25%. However there is a finite probability of there being zero from the yY cross, in this case with 20 offspring we obtain a probability of 0.003 that yY yields 0 melon offspring. Thus with twenty offspring from this cross we can be fairly certain if it is a YY or a yY. We will detail this analysis a bit further in this section.

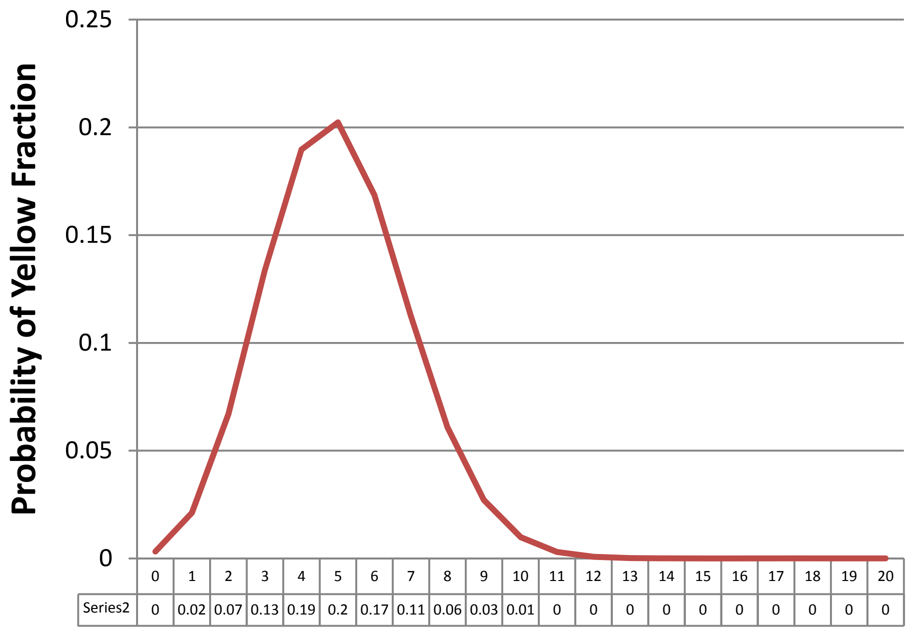


FIGURE 74 SELF CROSS OF YY X YY OR YY X YY

Now we want to look at the crossing of the parent with a melon, namely a fully recessive plant. First if the yellow is YY and we cross with a melon we obtain:

	y	y
Y	yY yellow	yY yellow
Y	yY yellow	yY yellow

This cross above says that if we were to cross melon with yellow and that if there were no yellows in the result then we could be certain that we had a YY as the melon.

If however we have a melon of the genotype yY then we have the following cross:

	y	y
Y	yY yellow	yY yellow
y	yy melon	yy melon

The result is that half of the offspring are yellow. Thus even one yellow yields a violation of the assumption of it being a pure melon. However one must also validate that the number of yellow offspring are in line with the assumption, namely we stipulate that there is a 1:1 relationship between yellow and melon. This means that we can stipulate two hypothesis:

Hypothesis 0 (H0): The yellow is YY and this means all the offspring are yellow.

Hypothesis 1 (H1): The yellow is yY and this means that half the offspring are yellow and half are melon.

Thus we want to perform a test to determine if the hypothesis 0 or 1 is true. However there may be a Hypothesis 2, namely none of the above. This means that we perform the cross and we obtain say 15% melon. What does this mean? It depends upon many factors, including the size of the sample. This is a classic hypothesis testing problem. We must then add a third hypothesis:

Hypothesis 2 (H2): None of the above.

Let us look a bit deeper into the analysis. If we calculate the fraction of offspring which are yellow we can define a variable as:

$$F_{Yellow} = \frac{Number_Yellow}{Total_Number}$$

But we know that this is a random variable. We know that if we cross yY with yy then there is a probability of 1/2 that it will be either melon or yellow. Then we know that if the cross is between yY and yy and we have n Yellow out of N samples, the probability that there are F yellow fraction is given by:

$$P\left[F_{Yellow} = \frac{n}{N}\right] = {}_{n}C_{N}\, p^{n}\left(1-p\right)^{N-n}$$

Which is the standard binomial distribution. Since p equals 1/2 we have:

$$P\left[F_{Yellow} = \frac{n}{N}\right] = {}_{n}C_{N}\left(\frac{1}{2}\right)^{N}$$

This is nothing more that the probability for a coin toss. As N gets large it looks like a Gaussian curve. The example below shows the results for a cross with 20 offspring. The probability that there are no yellows from this cross are 1 in a million. However the real question is what is the reliability that the model is itself true, namely that there is not

some other underlying probability, some other genetic mechanism that we are not observing.

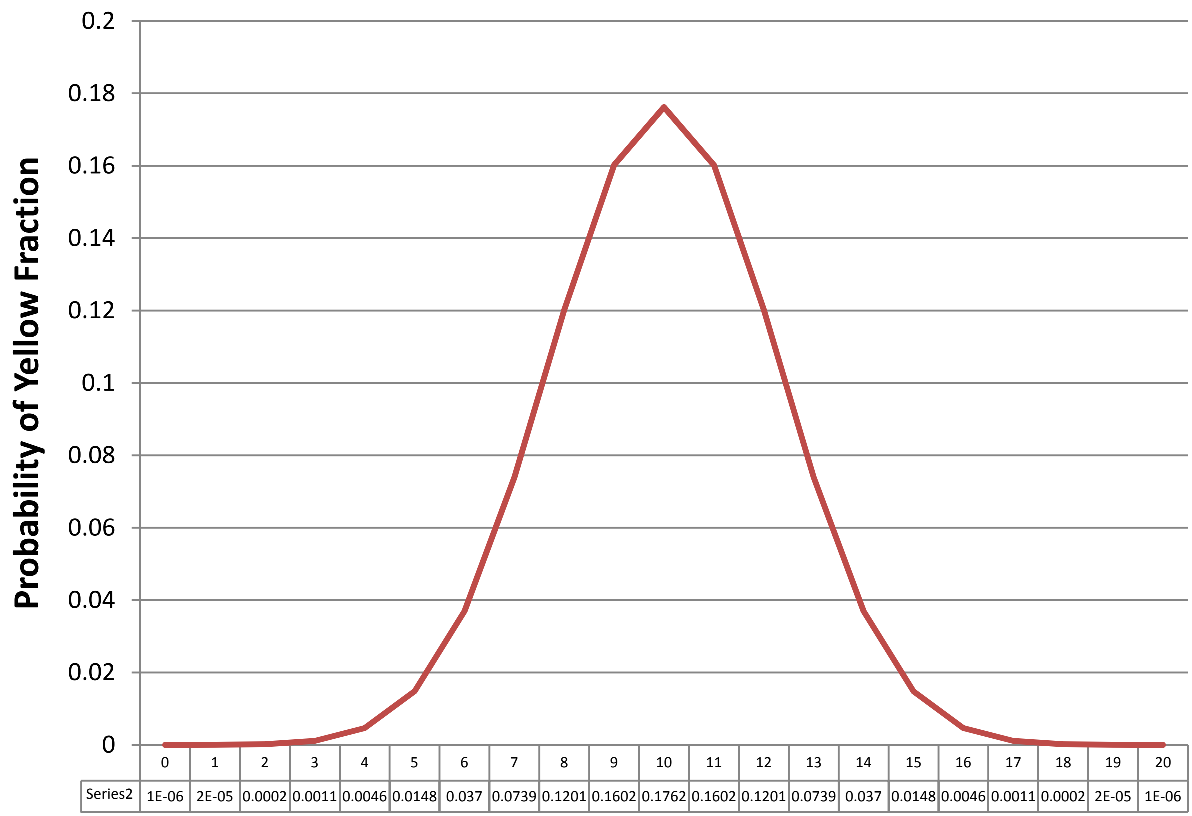


FIGURE 75 CROSS OF YY X YY AND YY X YY

This is the simple test that Norton should have performed.

7.5 SUMMARY OF MENDELLIAN APPROACH

We can summarize the world view of a Mendellian:

- Genes exists and are parts of a chromosome.
- There is a one to one relationship between a gene and some phenotypic characteristic. The genes control that characteristic.
- A gene may be dominant or recessive, namely there may be a stronger effecting gene than another.
- To get a characteristic the plant must have a gene which expresses that characteristic.

- Some genes are sex related or may have some effect on other genes but that is not a significant factor.

- The gene is the operative entity and there is not accounting for pathways, expression, activation or suppression.

- Mendel's approach fails to account for DNA and the underlying pathways.

The message to take away from the Mendellian analysis is simply; in hybridizing there is no simple one to one relationship between gene and phenotypic characteristic. What we see is a complicated system of variable gene expression; over and under expression, and the release of the gene products related thereto. We look at this in the next section.

7.5.1 Example

We will consider several simple examples of hybridizing.

7.5.1.1 Case 1: Hyperion and Species

The first is the hybridizing of Hyperion. This is a second generation from species and is shown in the next figure. The hybrid Florham was introduced in 1898 as one of the earliest hybrids. It is indeed a true hybrid being a cross between species H aurantiaca and H thunbergii. Florham seems to have been lost to history. In a similar manner the hybrid Sir Michael Foster is also lost. However, Hyperion is the result of Florham and Sir Michael Foster. Hyperion introduced in 1924 is still sold by multiple entities. It is in many ways one of the first commercial success for Hemerocallis hybridizing.

Hyperion

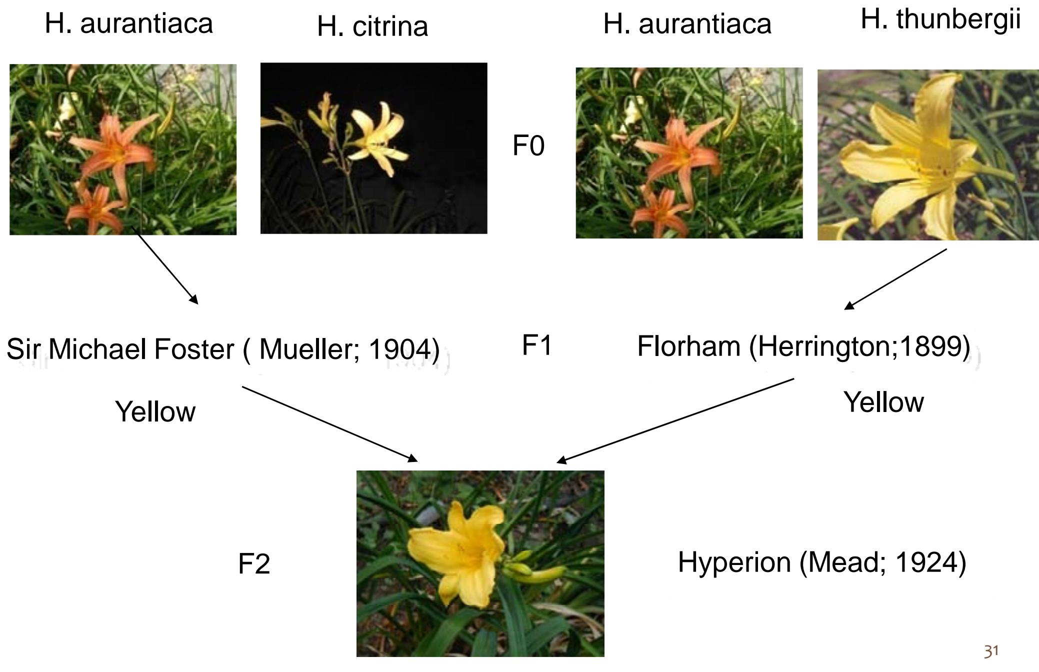


FIGURE 76 EARLY CROSS HYPERION

The above seems to suggest that at the F2 generation the yellow persists. However not seeing either Florham or Sir Michael Foster we cannot ascertain if the yellow was truly a recessive trait. In Stout's reference he states that the registration of the plant Florham states that the flower is a "canary yellow" thus seeming to infer that yellow is dominant in the thunbergii cross. Similarly in the same Stout reference we see that data on Sir Michael Foster indicates that it is also a clear yellow flower. Thus in this case we have a cross which may be:

	y	y
Y	yY Yellow	yY Yellow
Y	yY Yellow	yY Yellow

where y is the recessive of aurantiaca yielding the reddish color and Y is the dominant yellow color. This presumption is the antithesis of Norton. She argues for two separate genes, a yellow and a red. Thus she would say we have:
H aurantiaca: rR or RR. Also we would assume they are yy.

and for the F1 and F2 offspring as well as the F0 parents we have

H citrina and H thunbergii: yY or YY. Also we would assume that they are also rr. Let us do the crosses on these.

We have the following possibilities:

Case 1: rRyy X rryY

Case 2: RRyy X rrYY

Let us start with Case 1:

This case is rRyy X rryY.

	ry	rY	ry	rY
ry	rryy melon	rryY yellow	rryy melon	rryY yellow
ry	rryy melon	rryY yellow	rryy melon	rryY yellow
Ry	rRyy red	rRyY yellow (?)	rRyy red	rRyY yellow (?)
Ry	rRyy red	rRyY yellow (?)	rRyy red	rRyY yellow (?)

This yields the following result:

25% Red
25% melon
25% yellow
25% yellow (?)

Not having access to the detailed records we really cannot say at this time. However we know that the results chose were yellow and we have Hyperion upon which we can now experiment.

Let us now consider Case 2. This is for the cross RRyy X rrYY. This yields the following:

	rY	rY	rY	rY
Ry	rRyY yellow (?)	rRyY yellow (?)	rRyY yellow (?)	rRyY yellow (?)
Ry	rRyY yellow (?)	rRyY yellow (?)	rRyY yellow (?)	rRyY yellow (?)
Ry	rRyY yellow (?)	rRyY yellow (?)	rRyY yellow (?)	rRyY yellow (?)
Ry	rRyY yellow (?)	rRyY yellow (?)	rRyY yellow (?)	rRyY yellow (?)

If this were the case we would obtain 100% yellow (?). Perhaps this is the case. But we just as easily generate a dozen other likely profiles. Again the defects of Norton.

7.5.1.2 Case 2: Bicolors

The second example below is an interesting example of crossing with a bi-color. We started with Prairie Blue Eyes, since we desired to have the blue color. Then we crossed it with what was called Magic Dawn, but that name is in doubt, it was a bi-color[48]. We wanted blue and bi-color. These were the two characteristics we sought. The result was an F1 plant which was a non-descript red. It had no characteristic of either parent. Frequently this is common in the initial stages of hybridizing. There is a rule in hybridizing called the ruthless rule, where if a plant does not look good then get rid of it. Here we violated that rule.

We then crossed this with Karen Sue, a bi-color. From that cross came three name off-spring. Two of the plants below have a strong bi-color variation and one quite large and ruffled. These three now represent a based to further hybridize.

[48] From Terry Oates I was told that this may not be correct. See: http://davesgarden.com/guides/pf/go/18117/ Magic Dawn, Hybridized by Hall; Year of Registration or Introduction: 1954. The plant may be Howdy, see: http://davesgarden.com/guides/pf/go/26818/index.html Hybridized by Bremken-Armstrong; Year of Registration or Introduction: 1949.

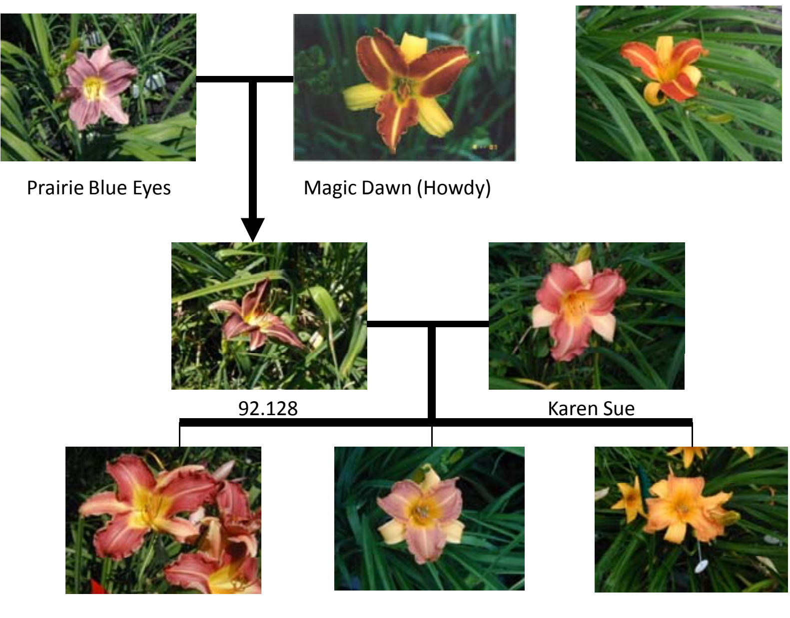


FIGURE 77 THE THREE SISTERS

Note, in neither case is there a compelling display of Mendellian genetics. We do in the second example see the persistence of the bi-color. However we generally get so few mature crosses that any good statistical results are not generally achievable.

7.6 Implications of Mendellian Crosses

In this section we look a bit more deeply as to the implications of the Mendellian method. We first discuss the concept of Heritability and then briefly introduce several of the classic techniques.

7.6.1 Heritability

Heritability is a concept in breeding which simply states that a certain characteristic or even characteristics which are phenotypical and which are quantitative rather than just qualitative have both a genetic and an environmental cause or influence.

Thus we look at the length of a scape, the width of a flower, the number of branches of a scape or even the total length of flowering as a quantitative element which can be measured. Then we say that this element or characteristic can be influenced by the underlying genetic factors and/or the environmental factors. It may be a hot summer, a dry summer, a clay field, a sandy field. All of these environmental factors may impact the measurement of the quantitative factor.

Now we can look at a factor, say the width of a flower, W, and we know that using this model we have:

$$W(i) = m + n_G(i) + n_E(i)$$

where m is the average width of this flower and the added factors are zero mean Gaussian variants with variances:

$$E[n_G(i)n_G(k)] = \sigma_G^2 \delta_{i,k}$$

$$E[n_E(i)n_E(k)] = \sigma_E^2 \delta_{i,k}$$

$$E[n_G(i)n_E(k)] = 0$$

$$\delta_{i,k} = \begin{cases} 1 \text{ if i=k} \\ 0 \text{ otherwise} \end{cases}$$

Then we define the total plant variance as:

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2$$

And the heritability is defined as:

$$h^2 = \frac{\sigma_G^2}{\sigma_P^2}$$

If h is greater than 0.5 we say that heritability is high for that characteristic and it is low if h is less than say 0.2. These of course are totally arbitrary values.

7.6.2 CREATING A HOMOZYGOUS LINE

Part of breeding program in the classic sense requires the creation of homozygous lines. Let us determine what must be done to obtain such a line.

Let us assume that a species is found in the wild. We do not know whether it is a homozygous, or dominant. There could be the following possibilities:

	Homozygous	*Heterozygous*
Dominant	*Case 1*	*Case 3*
Recessive	*Case 2*	*Case 4*

We now want to perform a set of crossing experiments to determine what we have. We start with the plant and self cross.

7.6.3 Heterozygosity and Dominance

Let us assume that we have any one of the four cases shown. We self cross and see what we can obtain. We do so assuming each of the four cases. Let us assume that the genes are T and t, for dominant and recessive.

Case 1: In this case we have TT for both. Thus we obtain:

	T	*T*
T	**TT**	**TT**
T	**TT**	**TT**

Case 2: In this case we have:

	t	*t*
t	**tt**	**tt**
t	**tt**	**tt**

Clearly we get the same phenotype in all crosses in both homozygous crosses and cannot tell what we really have.

Case 3: Here we have Heterozygous and dominant. This yields:

	T	*t*
T	**TT**	**Tt**
t	**Tt**	**tt**

But now we see e have three that look like the parent and one that is different and does not look like the parent.

Case 4: Now we have a Heterozygous and recessive. But this is impossible since if it is recessive it must be Homozygous.

Now we can say:

	Homozygous	*Heterozygous*
Dominant	*Case 1* *We obtain offspring looking the same*	*Case 3* *We get 3/4 looking the same and 1/4 looking different.*
Recessive	*Case 2* *We obtain offspring looking the same*	*Case 4* *Impossible case.*

Thus we have an ambiguity. Furthermore give a pure Homozygous of either a dominant or recessive we will never be able to tell. Thus we need two plants, of different colors, and from that we may have a better chance.

Now assume we have two plants, with two phenotypes, namely colors. Say a red and a yellow. We do not know which color is dominant and we do not know if the plants are Homozygous or Heterozygous.

Step 1: Self cross each plant to assess if the plant for each color is Heterozygous or Homozygous. We showed how this was done above. If there is more than one color we

know we have a Heterozygous plant and counting the frequency we can estimate the Dominant one.

If however we self cross and they both breed true to the same color as parents then we may have a dominant or recessive but each is Homozygous.

Now cross the two plants. We know one is recessive and one dominant. Thus we have:

	T	*t*
T	**TT**	**Tt**
t	**Tt**	**tt**

The above is an example which we had shown before. But we can now determine the dominant, since it is the dominant color.

From this experiment we first assess Heterozygosity and the second step we determine dominance.

7.6.4 Convergence of Homozygosity

Let us assume we have a plant which we know to be Heterozygous. We know that because when we cross it with a Homozygous recessive we get 50% of the recessive trait and we get a self cross with 25% of the recessive.

The question is how do we get a Homozygous Dominant plant? Simply we know that a cross of the presumptive Heterozygous plant with itself yields 25% Homozygous Dominant and 25% Homozygous Recessive. We want the 25% Dominant plants. So we get all of the Dominant plants, Heterozygous and Homozygous and do a test cross on the Recessive plant. If the results from a cross are all Dominant we know the parent is Dominant.

Let us assume we have a gene pair of Aa, and this is in the F0 generation. We now consider selfing or inbreeding in all generations. This means that the breeding is only with itself, no interbreeding. Thus by example we obtain:

F1, we obtain a cross of AA with itself, yielding AA, aa with itself yielding aa, and Aa with itself yielding AA:2Aa:aa. This means that of the 25% which were AA, they all breed true to AA, and likewise for the aa. But for the Aa which interbreed, and which represent

50% of the F1 population, they breed 25% AA, 25% aa and 50% Aa, thus we add another 12.5% to the AA and the same to the aa. This means we have only 25% which are Aa and the rest are equally split between AA and aa. We show the crossing in detail. All parents self cross. Thus at F1 the AA cross with AA and the aa with the aa. The same applies for all succeeding generations.

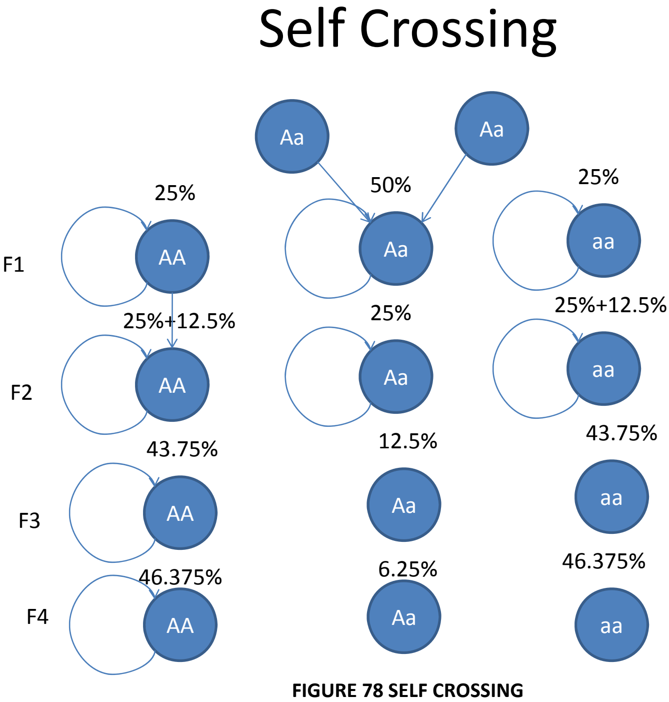


FIGURE 78 SELF CROSSING

F2, we again only allow self crossing. The same procedure results in the following Table. This can be continued and leads to the Table below.

Generation	*Genes*	*Homozygous %*	*Heterozygous %*
F0	Aa	0%	100%
F1	AA:2Aa:aa	50%	50%
F2	3AA:2Aa:3aa	75%	25%
F3	7AA:2Aa:7aa	87.5%	12.5%
F4	15AA:2Aa:15aa	93.75%	6.25%

Thus in almost no time we have bred homozygosity into the organism. The dominant are Homozygous and the recessive are by definition Homozygous.

Let us look at this in a bit more detail. In the Table below we show a Recurrent and non Recurrent. We start with a Recurrent with A genes and a non-Recurrent with a genes. At each descending generation we select as Fn the one with the non-recurrent at gene K

and we do not know what genes are at the other locations. However we always back cross with the A Recurrent but always select the a at gene k in the ensuing F state. The α value is a jth gene which we will analyze from the self crossing. We know the Recurrent is homozygous.

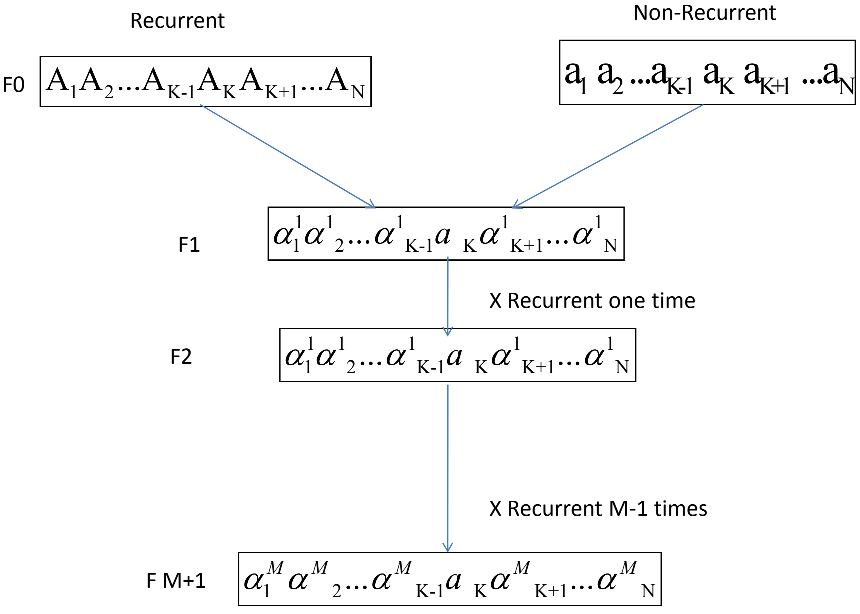


FIGURE 79 RECURRENT SELF CROSSING

First we assume the genes are independent and that we perform the self cross on the Recurrent. Now we can calculate the Recurrent percent per cross and this is shown below.

Crossings				
F0	AA		aa	
F1	AA 25%	Aa 50%	aa 25%	
F2	AA 25%	AA 25%	Aa 25%	Aa 25%
F3	AA 50%	AA 25%	Aa 25%	
F4	AA 75%	AA 12.5%	Aa 12.5%	
………				
FM	AA		Aa	
	$1-\frac{1}{2^{M-1}}$		$\frac{1}{2^{M-1}}$	

FIGURE 80 RECURRENT CROSSING

We note that after M crosses we have a percent Heterozygous of only:

$$\text{Fraction Heterozygous}=\frac{1}{2^{M-1}}$$

The fraction of Heterozygous becomes negligible as M increases. After say 11 crosses we will have less than 1 thousandth of the crosses being Heterozygous. The rate of convergence is quite fast.

7.7 HYBRIDIZING OR BREEDING TECHNIQUES

Before detailing some of the specific techniques, we will layout the process of setting goals and seeking the correct parentage to achieve those goals.

7.8 METHODS AND GOALS OF CROSSING

There are several classic mating methods. There are two dimensions in this process. The first dimension is choosing or selecting a plant. The second is how that selected plant's characteristics may be moved forward. Finally there is an algorithm for stopping.

To successfully develop a hybridizing technique a set of goals should be in mind from the outset. Here are a few examples.

1. Expanding Bicolor Flowers

3. Increasing Branching

3. Maximizing Bud Count

4. Extending Flowering Time

As we show above each of these has a quantitative measure. Thus we may start with plants that appear to take us on this path. We summarize this in the following Table.

Characteristic	*Quantitative Measure*	*Starting Plant*
Expanding Bi Color	Petal and Sepal color difference Petal and sepal colors	Use as source plants existing bi-colors. In addition select based upon pedigrees with consistent bi-color parents.
Increasing Branching		
Maximizing Bud Count		
Extending Flowering Time		

7.9 SELECTION METHODS

Let us begin by understanding the initial step, namely defining the goals and objectives to be achieved. There are several objectives in performing crosses in Hemerocallis. Several of them are:

1. To Generate an New Trait: This frequently comes about by pure random selection. If there were no spider to have as an example, then when one sees a flower with long narrow sepals and petals then this is a new trait and one may seek to both perpetuate it and to extend it. We may have no idea as to how this trait is controlled. This trait then can be in-bred many times seeking to extend the unique quality of the form. Thus we have seen more and more extreme variations of the spider, extremes in petal and sepal shape, variations in coloring, and variations in many other features while retaining the fundamental spider characteristic of a 4:1 or greater ratio of petal/sepal length to width.

2. To perpetuate and enhance a New Trait: Perpetuating a new trait, such as a bi-color, may require several generations of breeding, including multiple back crosses. The bi-color nature may be a recessive trait and the use of backcrossing with the original bi-color would re-enhance the bicolor nature.

3. To Modify an Existing Trait: We may like a bicolor or a particular eyezone and we may want to modify the flower to retail the characteristic while changing some specific color combination. We may want to keep the eyezone and blend the bicolor.

4. To Incorporate an Existing Trait: There may be a trait we want to corporate such a spider, bicolor, eyezone, or even just a simple color change.

5. To Test for Dominance of Traits:

7.10 Characterizing Goals

We must understand where we want to go and from whence we begin. There are several schools of thought that the hybridizer uses. But essentially they are divided into two branches.

First are those who take what is there and try to improve or enhance it. Thus many of the introductions are merely enhancements of what had been brought out before. For example, a ruffled flower with a contrasting eyezone and matching edges may be available as a new introduction. A hybridizer has a similar flower but in a contrasting color. The hybridizer then may try to do several additional crosses. First he may take the new hybrid and cross it with those of his own making, albeit not of the best color or form, and see what this new intro adds to his own collection. Or he may take the new hybrid and try to cross it with a flower of a color he is seeking is the more complex new hybrid.

Characteristic	*Hybridization by Extending*	*Goal Directed Hybrids*	*Targets of Opportunity*
Goals	Take next steps in introducing highly marketable plants.	Long term specific form and color goals. May include increased branching, viability, re-bloom, bud count and the like.	Seeking new and innovative features.
Initial Stock	Heavy use of third party hybridized stock for introducing new traits.		Heavy reliance of seeking out new and innovative internal hybridizing stock with specific features which are of interest and marketable.
Hybridizing Methods			
Data Keeping	This is almost a combination with Mass Selection and Pedigree. There is a keeping of records for parentage but the selection process is best of what was bred. Generally just use F1 offspring.	Requires extensive data and must keep records on all even those rejected. Photo records become a must in this area. F1 thru F6 generally are useful.	
Time Frames	May be the shortest of all because it builds upon already accepted introductions.	This is the longest process.	

7.11 Methods Applied to Crossing

There are several generic methods employed by hybridizers. In this section we present several of them as they may apply to Hemerocallis. Again as we has said before, the

goals intended should always be kept in mind. These techniques have certain advantages and disadvantages. In addition, many hybridizers look at "targets of opportunity", namely they look towards the "market" and what will sell at a particular time. In many ways this is typical of the general commercial horticultural market. This is unlike the agricultural market where the intent is generally one of seeking better yield, better protein content, better pest resistance and improved needs for fertilizer, water and the like. In the horticultural market it is an attempt to understand and follow the market trends.

I. this section we present an overview of some of the techniques. We include here certain methods which may be found more commonly in the agricultural area but in some ways may also have found their way into the world of Hemerocallis hybridizing.

We start with the broadly defined methods of pedigree and mass crossing. These methods are nothing more than on one extreme performing detailed crosses with the concomitant record keeping versus the method of just allowing "nature" to take its path and just select the best at each generation regardless of prior parentage.

There are generally two types of selection; pedigree and mass selection.

1. The pedigree selection method is a two step process. First, the plants pedigree, its parents and other lineage are tracked and recorded and this lineage becomes a factor in the choice of retaining and furthering the plant. Second, the phenotype is also a factor in the retention and furtherance of the plant. Pedigree selection is a selection process which attempts to balance the plants lineage and its appearance or other such usefulness.

2. The mass selection method is much simpler. At each step in the selection process, each generation, the best phenotypes are selected. An almost total disregard for lineage occurs in this process.

There are many methods of crossing and hybridizing and they can be performed in the context of either pedigree or mass selection. We will examine a few of the more classic ones in this section. Before doing so we examine the objectives of crossing. The techniques developed for agricultural plants and those used for ornamental plants share many of the same traits. We will not get into the details of the differences but will provide some detail on the many options available.

The following Figure compares Pedigree and Mass selection. In Mass Selection we just start with a cross, and then at each generation we select the plants we see as those having the best character at that generation. Namely we want a great deal of branching, then at each FN was to cross the plants with branching.

Let us first define the Pedigree and Mass Crossing methods in a general context. Both may apply to the hybridizing of Hemerocallis and we define them in further detail.

Pedigree: The Pedigree crossing method require the racking of plants parentage and using the characteristics of the parentage to pursue future generations. The reason for this approach is that we can often miss a recessive characteristic in F1 or even F2 and that if we want something from F0 we need to understand what F0 was and to hybridize to F3 or latter. Thus we must know the pedigree and this pedigree must be tracked and selection is based both upon the characteristic of the Fn plant as well as that of F0.

Mass Selection: If one looks just at each Fn and selects the desired characteristics from Fn independent of any prior generation, namely we could care less as to what parents we may have had, then we use that to propagate the next generation and repeat, we can see how Mass Selection works. It is especially good if we have a great deal of space and are willing to "just let nature take its course".

We compare these two below in an algorithmic form:

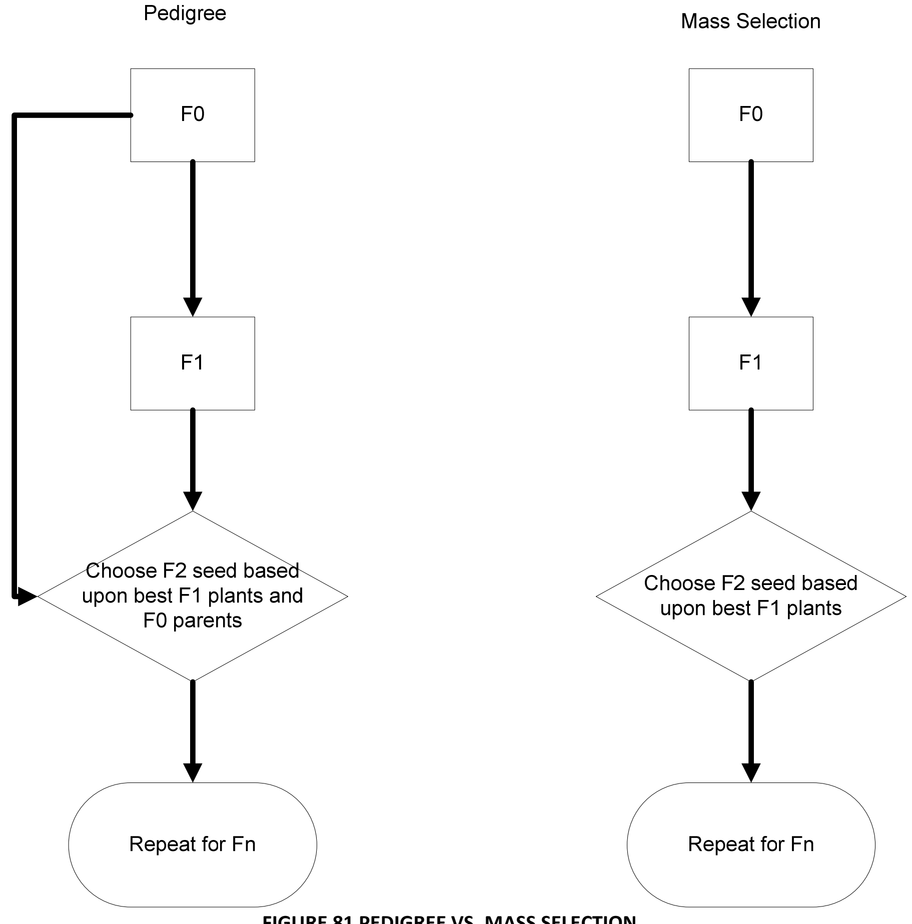


FIGURE 81 PEDIGREE VS. MASS SELECTION

The next question to ask is how pollination is performed. The various plants that one may see in a broad breeding mix are either self or cross pollinating. The Hemerocallis is a mix of both in the form of the species. However in hybridizing we generally hand pollinate and this means a cross pollination. However we may also desire to self pollinate to inbreed a specific trait as we had discussed earlier. Thus we can pollinate in one of two ways:

Self Pollinating: This means we take the plant with the most branching, or say the top ten such plants at F2, and we self cross them. This means we try to inbreed the characteristic in a line of plants.

Cross Pollinating: Here we use different plants, each with a large amount of branching and cross them.

We depict the various options in the following Figure.

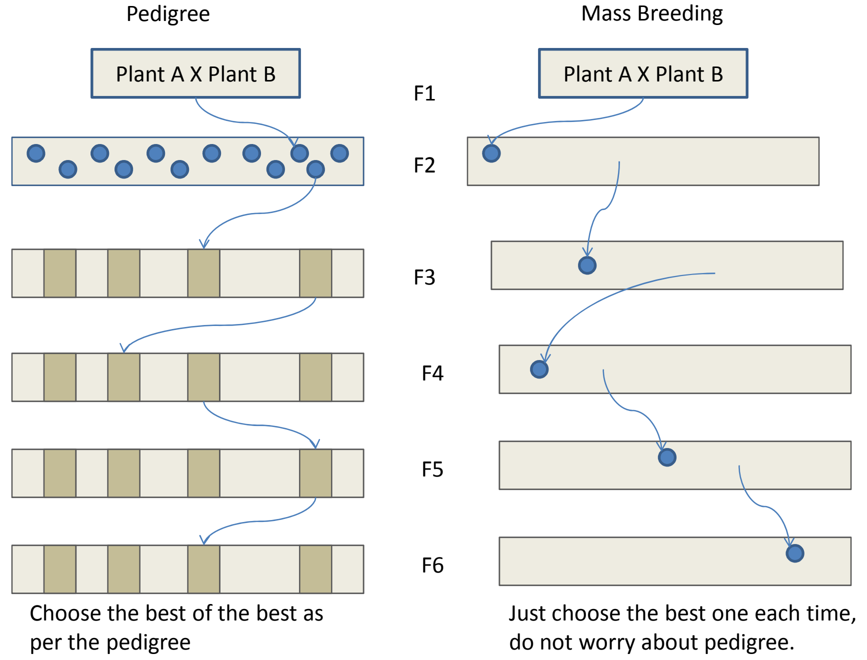


FIGURE 82 PEDIGREE VS. MASS OVER MULTIPLE GENERATIONS

We can now compare the four possible ways to proceed. We compare the two selection methods with the two pollination methods.

Pollination/Selection	***Pedigree***	***Mass***
Self	Choose best from a "family" at each round of selection. Self cross this collection of best in a family.	Select best in each round independent of any pedigree. Self cross and move forward.
Cross	Choose best from a "family" and cross the best from the same "family" intensifying the selected family trait.	Choose the best of the best independent of pedigree and cross these best plants.

7.12 CROSSING METHODS

We can now begin to examine the various crossing methods. These methods have been classified by Halinar and others are presented as follows[49]:

7.12.1 BACKCROSS

A backcross is a way to assess the parent who is dominant to determine if it is homozygous or heterozygous using the offspring. Namely we back cross the offspring onto the phenotypic parent. This we show below.

[49] See Halinar, J. C., Breeding Methods for Daylilies, The Daylily Journal, Spring 1990, Vol 45, No. 1, pp 24-30.

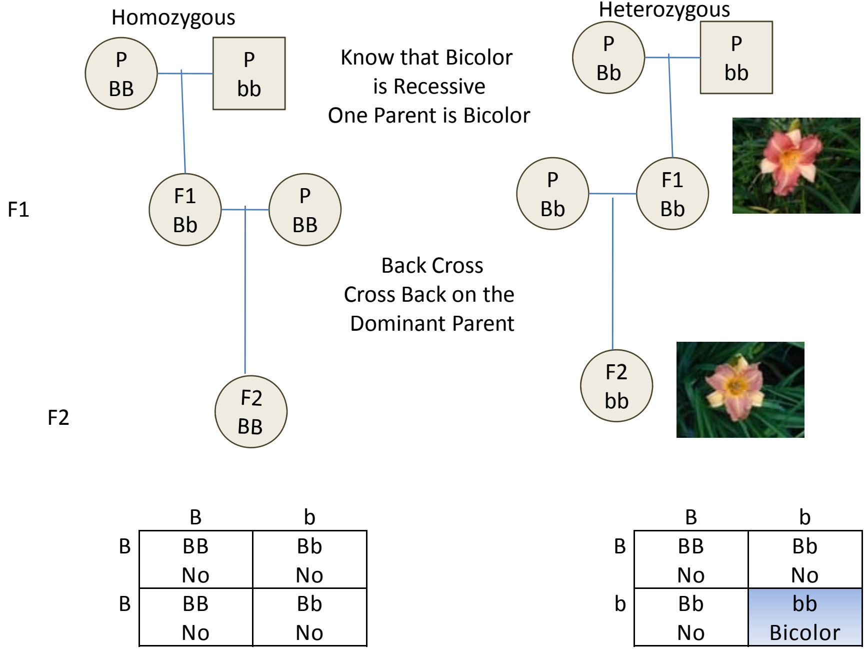


FIGURE 83 BACKCROSS

7.12.2 Testcross

The term Testcross has been used in a more general manner to describe crosses of putatively Homozygous dominant plants. We know that if we have a recessive gene being expressed in a plant then we must have all recessive genes and the plant must be homozygous. On the other hand the plant expressing the dominant characteristic may be homozygous in the dominant gene or heterozygous. We just cannot tell. Yet if we were to do a test cross between the two we would expect that any time we obtained a recessive phenotype using a recessive parent we have a heterozygous parent for the other plant.

The definition used in many works for Testcross details a great deal concerning its use:

"A Test cross is the mating of an incompletely known genotype to a genotype that is homozygous recessive at all loci under consideration. The phenotypes produced by a Testcross reveal the number of different games formed by the parental genotype under test.[50]"

7.12.3 Outcrossing

[50] See Stansfield Genetics p. 47.

Outcrossing is defined as the process of crossing cultivars or seedlings to unrelated cultivars or seedlings. The intent is to combine the characteristics of each parent into a sibling. This can work best with dominant traits which will end up in the cross. Thus was can cross Plant A with a dominant characteristic we desire with Plant B with another dominant characteristic and we hope to obtain a plant containing both characteristics.

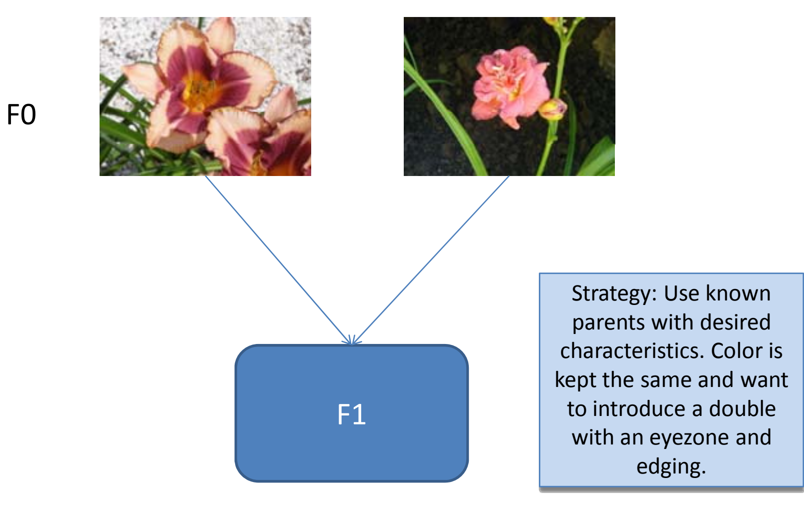


FIGURE 84 OUTCROSS

Thus outcrossing is simply taking two known parents with desired characteristics and trying to induce those characteristic in the offspring. As we have discussed elsewhere this yields an F1 generation where if there are dominant traits we shall see them but if what we are seeking is recessive we most likely will not. However, many hybridizers use this approach. It starts with a parent with desired characteristics and then crosses to enhance or expand that characteristic. We have seen this with eyezone flowers, ruffled edges, spiders and the like.

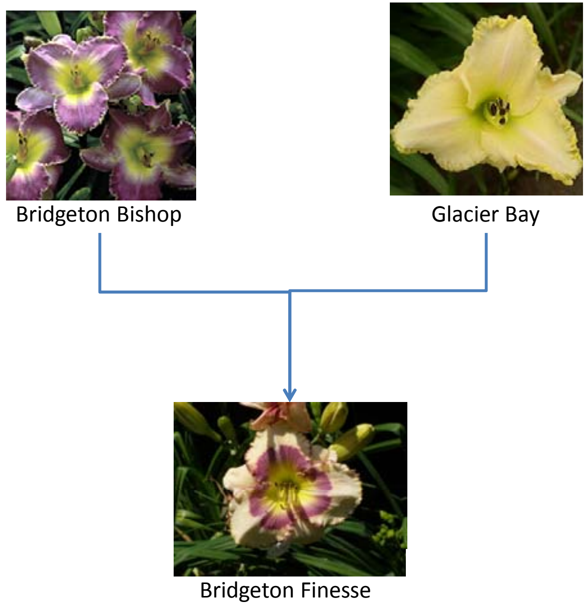
For example, using say a Kindly Light spider, one may cross it with other spiders to further extend the spider like characteristic. Typically the hybridizer stops at the F1 generation. The surprises however are all too often seen in the F1.

Some techniques of value in this area are:

1. Use of Known and Valued Parents: This means that many hybridizers have used the parent successfully and the new hybridizer will use this parent with other new parents for F1 results. Thus we may take a known recent introduction, which has been used by

other hybridizers for their new introductions and try to cross that parent with some of their own stock. Again it stops at F1.

2. Use of Identical Parents: This means that we try to duplicate the crosses that the originator had done. We may like a specific introduction and we may have its parents, if such are known. Then we can make the same cross again. The F1 results will most likely be different from the new introduction produced by the original hybridizer. However we may have a chance to extend the characteristics of the earlier introduction. Say we like Bridgeton Finesse. We know it parents are: Glacier Bay x Bridgeton Bishop. We get them and we cross them again. See the original cross below. We have a purple with a yellow throat crossed with a cream yellow. The F1 is an eyezone. Generally it is this unexpected result that is of interest.



3. Use of Similar or Substitution Parents: This means that we want say a ruffled eyezone. We know the parents of a desirable existing hybrid and we like the characteristics. However we will use parents of similar phenotypic characteristics.

We summarize these three variants of outcrossing below.

Type	*Characteristic*	*Advantage*	*Disadvantage*
Known and Valued Parents	Take parents with known and valued characters and attempt to create F1 with those characters combined or enhanced.	This may result in some new variation which has not yet been introduced.	There is no precedent for this type of cross. The chance that there may be a result is open to question.
Identical Parents	Take parents from selected existing hybrid with desired characteristics and redo the crosses.	There is a well established basis for the cross.	There may be a reason why there is only one introduction from the F0 of the original hybridizer. Also the closeness of the crosses may be so great as to make any new introduction valueless.
Similar Parents	Take parents with similar characteristics as those of a targeted existing hybrid and cross them.	Start with some basis for the end result.	

7.12.4 Line Breeding

Line breeding occurs when we cross related plants. Thus we can cross the plant with itself, its parent, its sibling, and its cousins. We have done that extensively in an attempt to obtain diverse bicolor characteristics where the bicolor is a recessive trait.

Line Breeding

F0

Use parent with desired characteristic and then cross to parent or other sibling. Attempt to induce a Homozygosity by inbreeding a characteristic.

F1

F2

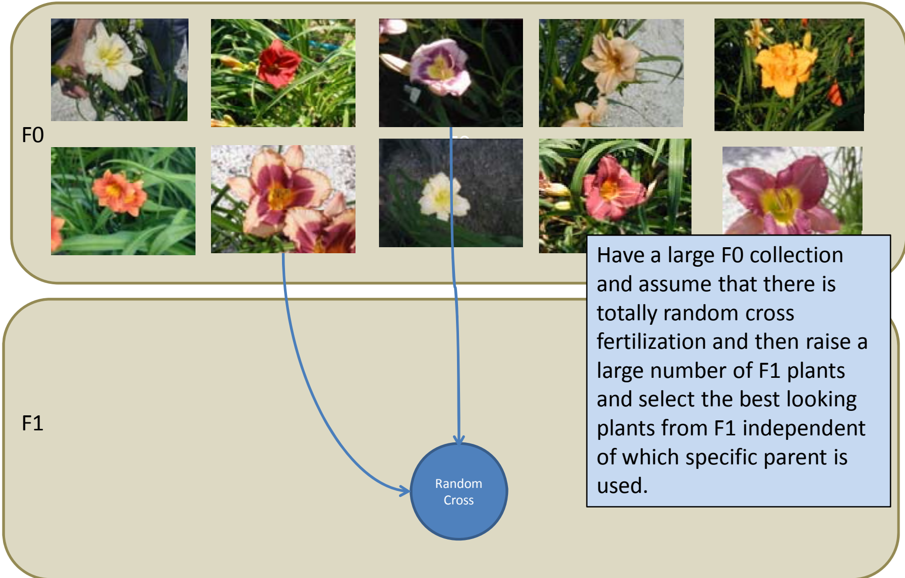
Line breeding is in many ways the most scientific. It does not end with F1 but can be continued. One may drive line breeding to the point of Homozygosity. The issue of closeness of parents is always a concern in line breeding and also the number of generations required.

In both Outcrossing and Line breeding we use knowledge of the parents and keep records accordingly.

7.12.5 Mass Selection

Mass selection we described earlier as a general principle as compared to Pedigree but as a specific methodology it is merely random fertilization of random plants and selecting the F1 descendents for the best traits. The base F0 parents are collected for the broadest possible set of characteristics and the F1 are selected based solely upon their phenotypic characteristics. This method is used rarely in Hemerocallis. It works well in a plant where cross pollination is strong and where there is plenty of room for selecting the F1.

Mass Selection

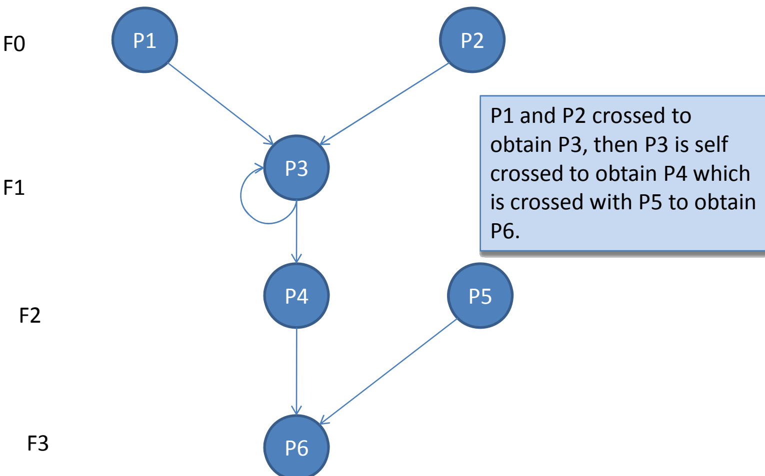


In many ways these are similar to the same methods used for crops in general which we have discussed.

7.12.6 Recurrent Selection

Recurrent Selection is a process which has the Pedigree methodology applied. In recurrent selection we take the F0 parents and then create an F1. Then the F1 is self crossed to yield an F2. Then the F2 is crossed with a selected parent. The objective of the self crossing to get F2 is to enhance the recessive characteristic. Say bi color is recessive. In F0 we may have one bicolor. Then, when we get the F1 generation, there are none. Then when we self the F1 to get F2 we may expect some bicolor again. We know this as a recessive, and Homozygous, and we cross that with a desired plant. We show this below.

Recurrent Selection



5. Backcross Sibling Mating: We start with general backcrossing. Backcrossing is the mating of an F1 with an F0 parent. It allows for the transference of a characteristic of one cultivar to another. Characters controlled by a single gene are readily passed on by this method.

Let us consider the following process:

(i) In F0 take two plants, one we shall call the Recurrent, and it is the plant we want to get a new characteristic into. For want of specificity we use H. multiflora as the Recurrent. We like the many branching and long blooms. We now want to get a bicolor trait into this plant. So we choose as a second F0 parent Howdy, a bicolor plant, see below.

Backcross

H multiflora: **Recurrent**

Howdy: **Non-Recurrent**

F1

F2

F3

To get F2, we back cross F1 with Recurrent, BUT we choose the F1 that has the character we are looking for in the Non-Recurrent. We repeat for several generations.

NOTE: All of the Fn generations we choose to back cross with the Recurrent have the phenotype of the Non Recurrent

(ii) In F1 we choose the plant which has the Non-Recurrent character, say bicolor, and we then cross it with the Recurrent parent, say H multiflora. This gives F2.

(iii) In F2 we again choose the plant with the bicolor and again cross it with H multiflora.

If we can assume that there is a single gene for this bicolor character, then we assume that we start with a Homozygous Recurrent and an unknown Non-Recurrent. There are two sets of genes we consider here. First, there is the single gene from the non-recurrent we want to transfer and second there are the many genes from the Recurrent we want to keep. In this case we want to transfer the bicolor gene from Howdy to the end result while keeping all of the M multiflora genes. Our goal is a bicolor multiflora.

Now we want to transfer the bicolor and we want to keep the rest of the multiflora. Let us assume all the genes are independent. Now phenotypically we have what we desire at any Fn if we select the bicolor. Yet we do not know if we have all the genes from the Recurrent H multiflora. How do we get to that point? Let us focus on a single gene, say A from the Recurrent and a from the non-Recurrent. We cross then generation:

AA X aa and this yields at F1, Aa and Aa. This means that at F1 we have mixed all the genes. By now backcrossing the selected plant at F1 with the Recurrent we obtain the cross:

AA X Aa and this yields AA:Aa and this means that 50% of the F2 plants are now Homozygous on the gene that was originally Homozygous in the F0 Recurrent parent.

If the gene that is transferred is dominant then we can do a self cross, namely what we have describe above in selfing, and we obtain a Homozygous result.

Backcross breeding thus works with the recurrent which breeds true from seeds. It allows for the introduction of a new trait.

7.13 BACKCROSSING: ANALYSIS AND STATISTICAL VALIDATION

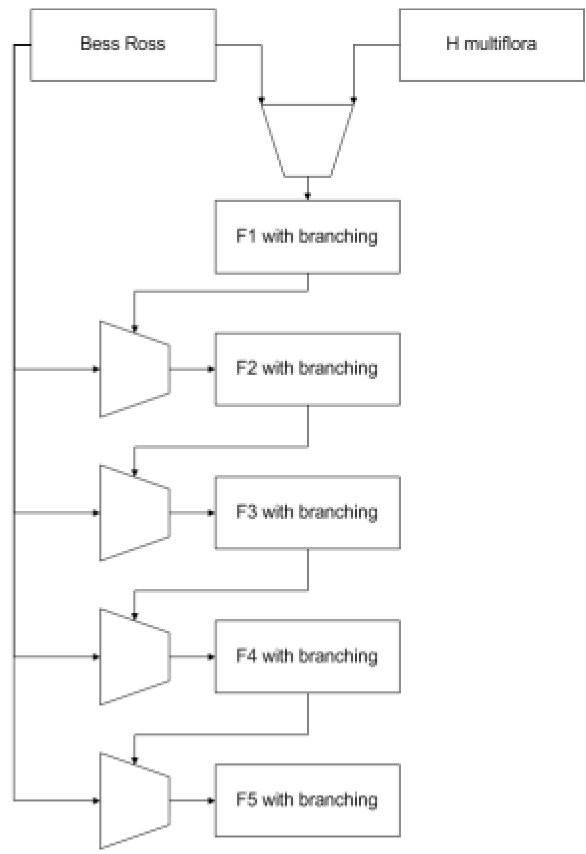
Backcrossing has been used for centuries. It is however frequently misunderstood and misapplied. In addition there appears to be limited mathematical models for the process of backcrossing and there thus results limited understanding of its application and capabilities. In this paper we review backcrossing using a specific Genus, Hemerocallis, and ten we develop a detailed mathematical model to analyze backcrossing in a generalized format. One of the key issues to be addressed is that of how many generations are required to assure an effective backcross, namely insertion of a desired gene, and the corollary question of how well this can be determined by a statistical analysis of the resulting backcrossed offspring. We also examining the inverse problem of estimating the number of operative genes which control the phenotypes based upon the measured results. Along with this problem we develop bounds on the accuracy of the estimation procedures.

Backcrossing is a simple process. One takes a plant with characteristics one is comfortable with, and then seeks to introduce a new characteristic from some other plant into the original one. For example, we may take the hybrid "Bess Ross", a diploid red daylily and seek to introduce into the plant a branching as one may find in the species H multiflora. We desire only the branching characteristic of H multiflora and we desire to retain all other characteristics of Bess Ross. The process we would employ would be backcrossing.

Backcrossing then works as follows. We first select a plant whose features we are satisfied with but for one characteristic. In our example we start with a diploid hybrid named Bess Ross, a red flower with no substantial branching. We want to introduce extensive branching into the plant. We want just the branching and not any of the other characteristics. Thus we say we desire to "drive" or insert the single characteristic of branching into the target plant. After the first cross, we then cross selected offspring, namely those with branching, with Bess Ross, again and again. After M such crosses we then ask what is the probability that we have the desired branched but otherwise homozygous Bess Ross. The result is then a plant which we could reproduce from seed and have a high level of confidence that it will breed true to form; namely a branched red flower appearing as a Bess Ross.

There has been an extensive amount written on backcrossing. The classic work of Allard uses a simplified two gene model and tries to exemplify the process. We argue herein that one must deal with the complex multi-gene model and no just two genes. The important issues result only when considering N genes. The recent work of Brown and Caligari also address the issue the same way. The results are frankly deceptive at best. The use of the approach in hybridizing horticultural plants requires a broader understanding of the issues. The work of Mayo also attempts to summarize the literature but we feel it too falls quit short of what is required. Brown et al also examine the issue but again do not address the details of the statistical model or the generalizations required. Similar high level analyses are performed by Griffiths et al as well as by Strickberger but failing in detail and depth.

The flow chart below depicts the details of standard backcrossing. It will be this process which we will analyze in some detail.



MATHEMATICAL MODELS

We start with the Recurrent plant, in this case the "Bess Ross" red diploid. It is assumed to have a collection of genes which control the flowering mechanism; These genes are assumed to control color, branching, budding, and the like. We assume that they act independently and are also on separate chromosomes and that further all plants have a homozygous form. Thus the Bess Ross genes are represented by the following dyadic. Each x is a gene and there are N such genes.

$$x_1 x_2 ....x_N$$
$$x_1 x_2 ....x_N$$

Now we have a similar gene for the species H. multiflora. There are also N controlling independent genes and the assumption of homozygosity again holds. Thus we can write a dyadic for the species as a collection of N y genes. This species plant from which we will seek to obtain the branching is called the Non Recurrent parent. It is shown below as a dyadic.

$$y_1 y_2 ....y_N$$
$$y_1 y_2 ....y_N$$

The desired outcome is a Bess Ross but with branching. We assume that branching is dominant. If it is not then we can obtain a recessive version readily by initially backcrossing with the H multiflora and then continue as we have stipulated. The target gene structure dyadic should be as follows.

$$y_1 x_2 ....x_N$$
$$y_1 x_2 ....x_N$$

The above endpoint is what we are seeking. Backcrossing will permit this to be achieved with a high statistical probability. Namely we would obtain after a selected number of crosses the Bess Ross but with branching.

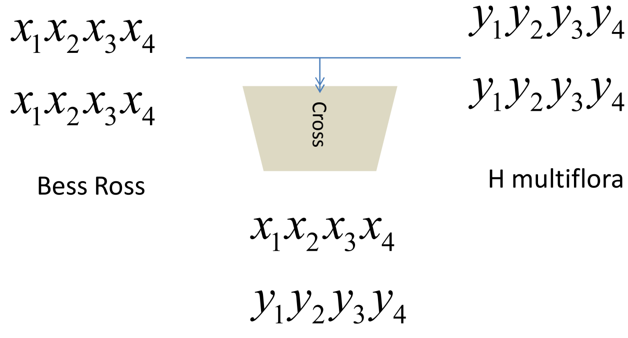
Consider a 4 Gene Case. Assume we want to insert y1 into the genome of the x sequence. Assume further that y1 is dominant. For example, we want branching from a H. multiflora to be placed into a red “Bess Ross”. The Example can be generalized to N genes and even M characteristics to be “driven: in from Non Recurrent into the Recurrent.

We start by crossing Bess Ross with H multiflora. All offspring will have the genetic makeup of the following dyadic:

$$x_1 x_2 ....x_N$$
$$y_1 y_2 ....y_N$$

The F1 generation is a pure mix of the genes from both parents. We shall assume that y1 is the gene for branching and that branching is dominant. If this is not the case then we can move to F2 by crossing with H multiflora and obtain a branched sample to begin the process. We assume that there are the M genes and that each gene results in a unique expression of some phenotypic characteristic which we can measure. We could assume that there is one for color and one for branching and neglect all others. This is the more classic approach. However as we have demonstrated before, we know that there are multiple genes required and that by allowing an unspecified pool of M genes that we can achieve significantly improved results. We demonstrate this first crossing below.

The F1 cross is as follows. All F1 are identical. We assume that both initial parents are homozygous. Namely they have identical genes on both chromosomes. We further assume that there is no linkage.



Copyright Telmarc Gardens 2008 All Rights Reserved 16

In the above we assume all genes are independent and not linked. The symbolic representation is just that, a symbol for the genes not their alignment. In fact they genes may likely be on different chromosomes.

$$Let\ X = \begin{Bmatrix} x_1x_2x_3x_4 \\ x_1x_2x_3x_4 \end{Bmatrix},$$

$$Y = \begin{Bmatrix} y_1y_2y_3y_4 \\ y_1y_2y_3y_4 \end{Bmatrix},$$

$$XY = \begin{Bmatrix} x_1x_2x_3x_4 \\ x_1x_2x_3x_4 \end{Bmatrix}$$

$$define$$

$$\tilde{X}_0 = \begin{Bmatrix} x_2x_3x_4 \\ x_2x_3x_4 \end{Bmatrix}$$

$$\tilde{X}_1 = \begin{Bmatrix} x_2x_3x_4 \\ y_2x_3x_4 \end{Bmatrix} or \begin{Bmatrix} x_2x_3x_4 \\ x_2y_3x_4 \end{Bmatrix} or \begin{Bmatrix} x_2x_3x_4 \\ x_2x_3y_4 \end{Bmatrix}$$

$$\tilde{X}_2 = \begin{Bmatrix} x_2x_3x_4 \\ y_2y_3x_4 \end{Bmatrix} or \begin{Bmatrix} x_2x_3x_4 \\ y_2x_3y_4 \end{Bmatrix} or \begin{Bmatrix} x_2x_3x_4 \\ x_2y_3y_4 \end{Bmatrix}$$

$$\tilde{X}_3 = \begin{Bmatrix} x_2x_3x_4 \\ y_2y_3y_4 \end{Bmatrix}$$

Note that genes xn and yn are equally likely and have probability ½. Note that if we look at the gene tails, if they are M in length then we have $[1/2]^M$ for any one of them. Note further that for the combinations of 0, 1, 2, 3, etc we have the binomial distribution to provide the probability for any possible set of transitions from one F generation to the next F generation.

The following is a set of such transitions which are possible for this specific example. It should be readily determined what the transitions would be for any generalized form. The notation can be described as follows. If we have a cross between X0 and X0 then we can only get X0. Is we have a cross between X0 and X1, where this means that we have a tail sequence with just one y gene amongst the group, then we can get either an X0 or an X1 with equal probability. The same can then be said if we have an X0 crossed with an X2, yielding an X0, or an X1, or an X2, but now the result is controlled by a binomial distribution. The process than continues. We show the results with a three independent gene tail as follows:

$$X_0 \oplus X_0 = \{X_0$$

$$X_0 \oplus X_1 = \begin{cases} X_0; \textit{with probability 1/2} \\ X_1; \textit{with probability 1/2} \end{cases}$$

$$X_0 \oplus X_2 = \begin{cases} X_0; \textit{with proability 1/4} \\ X_1; \textit{with probability 1/2} \\ X_2; \textit{ with probability 1/4} \end{cases}$$

$$X_0 \oplus X_3 = \begin{cases} X_0; \textit{ with probability 1/8} \\ X_1; \textit{ with probability 3/8} \\ X_2; \textit{ with probability 3/8} \\ X_3; \textit{ with probability 1/8} \end{cases}$$

Now we can consider the transition from F2 to F3. Recall that F1 is merely a set of genes sharing one from each parent, the x,y combination. Then for F2, which is F1 crossed with the all X parent, we have the first form of segregation, namely we can get as the three gene tail, an all x, a one y and two x, a two y and one x, and a three y set.

To perform this analysis with a three gene tail, we will perform the analysis for each possible combination. We create a Table which shows what the crossing gene sequence is, say an X0, X1 and the like, and we then show a column which is the probability of that sequence in F2 and then we have a column for the transition probability of that sequence in F2 to the X0 sequence in F3, or the X1 sequence in F3 and so forth. This is shown below first for the X0 transition and then all others:

Cross	*Prob of This Cross in F2*	*Prob of X0 in this Cross*	*Prob X0 at F3*
X0	1/8	1	1/8
X1	3/8	½	3/16
X2	3/8	1/4	3/32
X3	1/8	1/8	1/64
Total Prob X0 in F3			27/64

Now we perform the analysis for the X1 cross elements. The second column remains the same but the third column reflects what we had demonstrated earlier. If the tail is X0 there is no chance of getting an X1 since there would be no ys available. Likewise for the X1, X2, X3 crosses we would expect a reduced number of corresponding tails in the ensuing generations.

Cross	*Prob of This Cross in F2*	*Prob of X1 in this Cross*	*Prob X1 at F3*
X0	1/8	0	0
X1	3/8	½	3/16
X2	3/8	1/2	3/16
X3	1/8	3/8	3/64
Total Prob X1 in F3			27/64

As we move to the X2 and then X3 we see that the number of them decreases at a faster rate as shown in the table below.

Cross	*Prob of This Cross in F2*	*Prob of X2 in this Cross*	*Prob X2 at F3*
X0	1/8	0	0
X1	3/8	0	0
X2	3/8	1/4	3/32
X3	1/8	3/8	3/64
Total Prob X2 in F3			9/64

Finally for X3, we see that only the tail in X3 of the prior generation do we get the chance for an X3, and that gets smaller geometrically each additional cross.

Cross	*Prob of This Cross in F2*	*Prob of X3 in this Cross*	*Prob X3 at F3*
X0	1/8	0	0
X1	3/8	0	0
X2	3/8	0	0
X3	1/8	1/8	1/64
Total Prob X3 in F3			1/64

Note that the second column is the probability of the specific sequence in F2 and that the third column is the transition probability at that specific cross to the next F generation. Namely the third column is the probability:

$$P\left[X_k(F_{n+1}) \mid X_j(F_n)\right] = p_{k,j}(n)$$

and

$$P(n) = \begin{bmatrix} p_{0,0} \cdots p_{0,N} \\ p_{N,0} \cdots p_{N,N} \end{bmatrix}$$

The above are the transition probabilities and can be readily shown to be independent of the specific crossing state, namely which Fn the probability of made for. Now we can calculate the probability of any Xn for a specific state Fk. This is as follows:

$$P[X_n(F_{k+1})] = \sum_{i=0}^{N} P[X_n(F_{k+1}) \mid X_i(F_k)]P[X_i(F_k)]$$

We have shown above that the transition probabilities are state independent and that the above equation is a recursive means to determine the next state. We demonstrate this for F4 from F3 as below:

We now do F4, and again we select the plants expressing Y1 and we again back cross with the homozygous X. This follows the same logic we did for F3. This then yields a 67% Homozygous for F4 with three genes other than the one we want impressed. The Table above can then be iterated again and again. We simply use 342/512 in the second column.

Cross	*Prob of This Cross in F3*	*Prob of X0 in this Cross*	*Prob X0 at F4*
X0	27/64	1	27/64
X1	27/64	½	27/128
X2	9/64	1/4	9/256
X3	1/64	1/8	1/512
Total Prob X0 in F4			343/512= 0.67

Cross	*Prob of This Cross in F3*	*Prob of X1 in this Cross*	*Prob X1 at F4*
X0	27/64	0	0
X1	27/64	½	27/128
X2	9/64	½	9/128
X3	1/64	3/8	3/512
Total Prob X1 in F4			147/512= 0.287

Cross	*Prob of This Cross in F3*	*Prob of X2 in this Cross*	*Prob X2 at F4*
X0	27/64	0	0
X1	27/64	0	0
X2	9/64	1/4	18/512
X3	1/64	3/8	3/512
Total Prob X2 in F4			21/512= 0.041

Cross	*Prob of This Cross in F3*	*Prob of X3 in this Cross*	*Prob X3 at F4*
X0	27/64	0	0
X1	23/64	0	0
X2	9/64	0	0
X3	1/64	1/8	1/512
Total Prob X3 in F4			1/512

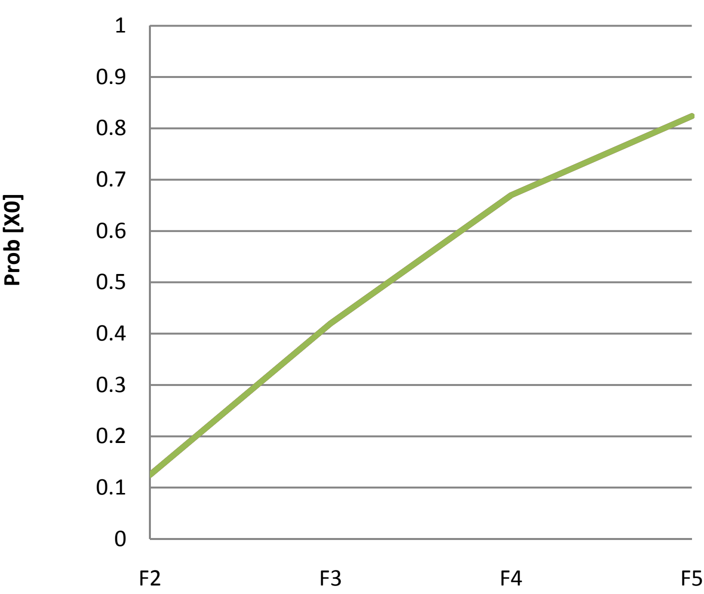
Finally we can extend this one further time to the F5 from F4 states, focusing solely on X0. This yields the following Table using the models developed above:

Cross	*Prob of This Cross in F4*	*Prob of X0 in this Cross*	*Prob X0 at F5*
X0	0.670	1	0.670
X1	0.287	½	0.144
X2	0.041	1/4	0.010
X3	0.002	1/8	0.000
Total Prob X0 in F5			0.824

We now have a simple algorithm: The column for the last cross must be iteratively calculated for every prior step as shown. The column for the probability at the current cross can be calculated once, they will be binomial in form. The probabilities for the

current and then next cross can be calculated by summing the products. Note that the larger the genome in the Recurrent the more complex and the longer the convergence.

Then we can plot the convergence rate to homozygosity in the graph shown below. Note that at F5 we have gotten to 82.4% of homozygosity.



Analyses for more complex genes and for more lengthened crossings can be accomplished. However the principle is shown in the above example. The key point to make is that the analysis we have performed herein is essential more realistic than the simplistic ones performed in the literature.

7.14 Statistical Analyses

There are many statistical issues relating to this analysis. In this paper we focus primarily upon two issues.

First, if we assume we know M, the number of controlling genes, and we know that the model is correct, then we can determine how many crosses, N, will be required to obtain a level of selection as may be desired. One way to validate this is by testing the means of the various clusters that result and determining if they are converging at the required rate. We develop a simple test to verify this and establish bounds on the results.

Second, there is the issue of estimating the number of controlling genes, N, that may be in the backcrosses. This is a corollary to the first problem. Namely if we have two plants, each with a certain number of distinct phenotypic characteristics and we assume that we have one gene and one phenotype, then the question is how many genes are in this backcross mix? We have assumed that we know N, the number of genes. In reality we most likely do not know N, however we know the number of generations by definition, we have measures on the phenotype characteristics and their respective frequency. Thus we should have enough to obtain an estimate of N by using the assume convergence model developed herein, and furthermore we can obtain bounds on the accuracy of the estimate of the value of N obtained thereby.

We first consider the question of how many generations we must cross to attain a desired level of homozygosity. We know from classic t-statistics how to size and experiment for a specified level of certainty if we were to see if the mean were within certain bounds and within the desired level of certainty. There are also simple tests to determined paired samples. However in this case the problem can be stated more complexly. We have N characteristics and we know what the means are for the number of samples in each of the characteristic sets. We further know that as we increase the number of crosses M to a larger number that the average number in the sets being crossed against decrease exponentially. In reality we only desire to retain the set for which we are backcrossing and whose presence is exponentially increasing. Thus the determination of the number of samples required to reach a level of confidence may be obtained by focusing on the X0 set only and then doing so in each Fn generation (see Pagano and Gauvreau).

We can now address the second issue. Namely, given a set of sequential measurements of phenotypes, what is a reasonable estimator of M, the number of genes controlling the phenotypes. Consider the following experiment. Let n be the nth cross, with corresponding generation Fn. Let there be a total of N such generations. Let B be any resulting set of normalized results for a phenotype in that generation. We will detail this as follows:

$$B_k(n) = \frac{T_k(n)}{\sum_{i=1}^{M} T_i(n)}$$

where $T_k(n)$ *is the total with phenotype i at Fn*

Now we know that:

$$P[X_k(n+1)] = \sum_{j=0}^{M} p_{k,j} P[X_j(n)]$$

Which we can write as:

$$T_k(n) = T(n)P[X_k(n)]$$
$$\textit{and T(n) is the total number}$$
$$\textit{in the Fn generation}$$

Now we can also look at each of the values of T or equivalently the normalized values we define as B, as follows.

$$P[M \mid B] = \frac{P[B \mid M]P[M]}{P[B]}$$
$$where;$$
$$B = \{B_0(1)....B_M(N)\}$$

But we also can say that:

$$B_k(n) = \overline{B}'_k(n) + w_k(n)$$
$$where$$
$$\overline{B}'_k(n) = P[X_k(n)]$$

We can use a maximum likelihood estimator which gives M as follows:

$$\textit{Find M to maximize:}$$
$$P[B|M] =$$
$$P[B_0(1)...B_M(1)...B_0(N)....B_M N) \mid M]$$

Now we can use the previous observation to state that the Bs have known means, given M, and that we can calculate them, and that they are random variables with w being a zero mean Gaussian with variance σ and we can further assume that they are independent. Then using the log of the likelihood function as defined we can then obtain an estimator which minimizes that sum of squares. Now we need to determine the variances on each of the samples. The variances will be used to weight each sample. Before proceeding we can restate the ML solution as follows:

Find M to minimize:

$$\sum_{n=1}^{N}\sum_{m=0}^{M}\frac{(\overline{B}_m(n)-B_m(n))^2}{\sigma_m^2(n)}$$

We can use the sample variances for the ensemble variances. Similarly we can calculate the ensemble variances using the fact that the ensembles are generated by the binomial selection processes. The ensemble variances are quite difficult to calculate so we retain the sample variances as simpler measures.

Now we can determine the variance on the estimate by using the Cramer-Rao bound which functions well on such Gaussian analyses (see Van Trees). Specifically we have:

$$\mathrm{var}(M-\widehat{M})\geq\left(E\left\{\left[\frac{\partial^2\ln p(B\,|\,M)}{\partial M^2}\right]\right\}^{-1}\right)$$

But since these variables are assumed Gaussian this can be calculated readily for any M.

As an example, we could consider the crosses we had discussed above. If we look at Bess Ross and H multiflora, we could consider 2 genes, color and branching, and then go from there. For three genes, we could introduce the root, tubular versus bulbous, then length of scape, length of leaf, width of leaf, number of flowers per branch, and so forth. We note that as we increase the number of putative genes, the denominator which represents the total number of samples, goes up, driving the ratios for each gene down. As we increase the genes we then get more variation and it goes up again. Thus, arguably there is a minimum.

The method proposed is actually a form of cluster analysis (see Fukunaga). It seeks to find the optimal number of clusters of values for sets of characteristics. By examining the method, the clusters are based upon a collection of characters. For example, if we have N=2, then for all branched plants we have color and scape length as possible characters. We then sort on the four possible sets; red and long scape, red and short scape, yellow and long scape and yellow and short scape. The Bess Ross could be defined as red and short scaped. We could then also expand it to the other characteristics as we have discussed before.

7.15 Discussion

The ability to backcross is an essential element in hybridizing. It permits the introduction of a trait into an existing line and then ensuring that the line is returned to its original genetic state with the exception of the new phenotypic characteristic having

been expressed. All other phenotypic characteristics are returned to where they were at the initial state.

There are several additional enhancements which must be made to this analysis. First, linkages must be incorporated. For F1 through typically F5 the linkages of genes may not play a significant role. However as we continue to backcross there are increasingly import effects of linkages which must be accounted for (see Griffiths). Second, we know that many of the genes are modulated by repressor and activator genes. These must also somehow be accounted for. Generally, if they are not affecting other genes we can let them be second order effects. However, when they cross modulate in gene expression motifs then we have to establish their presence in the model. Third, this is an analysis and hybridizing planning tool. This is not a synthesis tool as currently structured.

7.16 Comparison of Methods

We now compare them in the Table that follows:

Type	*Characteristic*	*Advantages*	*Disadvantages*
Outcrossing	Crossing with parents with desired traits in an attempt to combine.	Potential for selecting traits to be combined or carried forward. Can identify a trait in a parent phenotypically.	Single outcrosses may be controlled by dominance or even hidden traits which are not transferrable in one cross.
Line Breeding	Use self crossing to attain recessive traits. May also cross on close relatives like first degree siblings to enhance the trait.	Allows for the use of genetic principles to force a trait into a line. May be able to use a recessive trait by selfing or close crossing.	May take a long time due to multiple generations.
Mass Selection	Start F0 with large selection of plants with good characteristics. Allow random mating. Then grow F1 and select best phenotypes from there, regardless of parent. Continue the process.	Ease of implementation and no need for records. Just cross and select.	Lacks controllability and total lack of selective breeding techniques.
Recurrent Selection	In F0 use two parents with desire traits. In F1 self cross to enhance any recessive trait that may have been hidden to obtain an F2. In F2 cross with another desired trait to get F3 which is the generation for choosing.	Useful for enhancing a recessive but desirable trait.	Quite complex and tedious and takes a great deal of time. If it takes at least two years for each generation we may require easily six years to get to the first selection point.

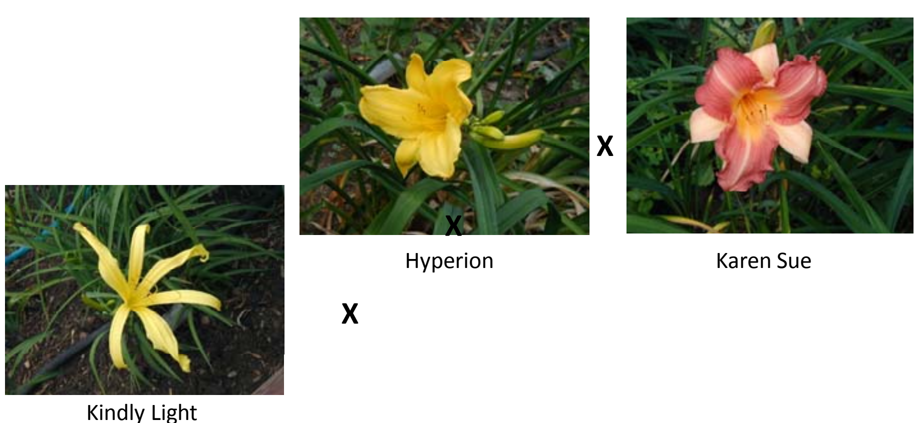
Backcross Sibling Mating	Select a Homozygous plant whose characters you want to keep except for say one. Select another plant with that character. The first is the Recurrent and the second the non-Recurrent. Cross them and select the one with the desired new characteristic. Cross that with the Recurrent, again select the one with the characteristic and repeat.	It is possible to place a new characteristic into an existing Homozygous line while keeping the rest homozygous. This is a genetically based approach.	Takes a great deal of time and many generations.

7.17 Hybridizing Examples

In this section we present some examples of hybridizing we have been involved in and explain the logic used to obtain the end results presented.

7.18 Bi-Color and Spider

The first example is shown below. We started with Hyperion and Karen Sue. Our objective was to use the vigor of Hyperion, including the branching and the strong flowering and introduce the bi-color of Karen Sue. Based upon our prior work we see that perhaps the bi-color is recessive. It did not appear in the parent.



2 Offspring from the same sets of crosses. Note that Grant is no true bicolor whereras Hilda is.

Happy Hilda

USS Albert W Grant

FIGURE 85 BICOLOR PARENT AND SPIDER, TWO DAUGHTER PLANTS

7.18.1 F2 Bicolor

In a similar cross we crossed Hyperion X Karen Sue with itself and obtained Rita's Sunrise. A large multi-branched yellow flower with a red eyezone. This is shown below. It has Hyperion as a base, but is much larger apparently getting size and more from Karen Sue, it is not bicolor as is Karen Sue but it has apparently picked up the red from Karen Sue and it is displayed in the eyezone.

Rita’s Sunrise

FIGURE 86 F2 CROSS

When looking at the above cross one may ask what the objective was. Simply we chose Hyperion for it extensive branching and bud count. It provided vigor. We chose Karen Sue for the bicolor nature. We were trying to obtain a bicolor which had vigor. Rita's sunrise was a plant which has the vigor but not a bicolor yet it has a very assertive eyezone. It has the yellow of Hyperion, again saying that yellow seems to be dominant, and the red of the Karen Sue is carried only in the eye. Rita's Sunrise was an F2 cross of Hyperion X Karen Sue.

7.18.2 F2-F3 Eyezones

Now let us look at a more complex cross as shown below. Originally we tried Roy Beaver a yellow aggressive growing plant with Prairie Blue Eyes. The objective was to try to get the blue in the yellow. This was before we understood the dominance of yellow. Then we crossed it with Wine Bold to see if we could obtain the red to suppress the yellow dominance. Clearly the color of Prairie Blue Eyes is driven by the red not the yellow. In another cross we did Royal Kingdom and Whoperee since we wanted an eyezone and a dark red. This result of a set of crossings led to Bishop Gabriel. It is a reddish flower with a large inflorescence, good branching and a throat. At best it looks somewhat like Wine Bold but is bigger and more branched.

From this crossing we can learn the following:

1. Dominance of traits must be understood, They will control the results of many crosses.

2. Recessive traits like the red color can be brought out but it takes several generations.

3. Goals can be flexible. Our original intent was blue. This was clearly not met, nor frankly has any hybridizer met the goal. For example one would logically think that crossing Prairie Blue Eyes with Prairie Moon, an almost white daylily may carry over the blue color into a white plant. However the white is a dominant color over the blue.

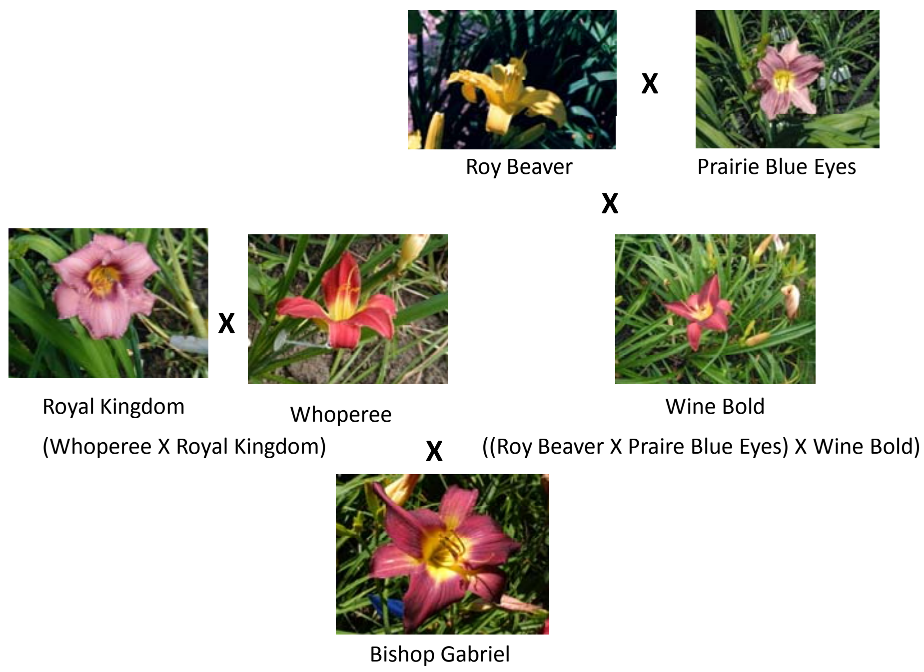


FIGURE 87 BLENDING OF COLOR, SIZE AND FORM.

7.18.3 F2 Eyezone and Color Change

The next cross shown below is also surprising. We crossed Cynthia Paige Platais with Love Festival. In this case we were seeking reds with eyezones. Out came Princess Martina. Princess Martin is a yellow flower with a red eyezone. Again we see the dominance of yellow. Even though neither parent had all yellow, at best they both had yellow throats, Princess Martina kept the yellow throat and the ends of the petals and sepals were turned yellow.

Cynthia Paige Platais

X

Love Festival

Princess Martina

FIGURE 88 DRAMATIC CHANGE IN DAUGHTER

The above further demonstrates the yellow dominance. In a strange manner the flower retains yellow at the ends and in the throat. The red becomes the remnant rather than the dominant factor. As we have discussed before in the analysis of color we still have the intriguing issue of color variability across the flower.

7.19 BICOLOR AND DOMINANCE

Consider now the following cross. We used Karen Sue with American Belle. The goal was clearly a bicolor with American Belle as almost a background color. The resulting cross, Sara's Dreams is a dramatic shift again. We have an orange red, with a yellow green throat and re-curved petals and sepals.

X

American Belle

Karen Sue

Sara's Dreams

FIGURE 89 UNEXPECTED DAUGHTER

7.20 Blending versus Dominance

The following is another example of a cross where the result is mixed. If we were to return to the Norton model, here we have a classic case of a red and a yellow. We studied just this case in the last section and agreed that if Norton were correct yellow would dominate. However as we look at this result it is a blend! It is a purple flower and there is no evidence of a red or a yellow. Thus the simple genetic dominance theory proposed by Norton seems not to hold here at all.

The cross was Superchild with Love Festival. The intent was to use Superchild as the base for a large tall flower and use Love Festival to gain a red Superchild. The result was Maja's Tinkerbell. It is a pastel purple flower with white ribs on the petals and a green yellow throat. It has the strength of Superchild but is more akin to an off-spring of a Prairie Blue Eyes.

Love Festival

X

Superchild

Maja's Tinkerbell

FIGURE 90 DIVERSITY OF COLOR, CLEAR EXAMPLE OF MIXING

The above examples show the diversity of results and supply limited knowledge of the true genetic makeup of the plants.

7.21 CONCLUSIONS

The process of hybridizing is a bit science and a bit art, it is a bit strategy and a bit whimsy. We have summarized the classic Mendellian approach and then we have reviewed the classic methods or breeding as understood under the Mendellian rubric. What we see is that the hybridizing of the Hemerocallis is often less the rigorous approach taken by those who breed for crops and is a hit and miss affair, with some idea of where they are going.

We have seen distilled certain rules of hybridizing:

1. Start with good stock. This is obvious in Stamile. They have breed their own good stock and they then select the best of the best. The same is true of Petit. In contrast Mahieu blends good stock with species, specifically H citrina. It all depends on what one views as intent but we see Mahieu as a leading edge innovator bringing back characteristics that may all to easily be lost in the rush to the extreme.

2. Use your own innovations. If a hybridizer has talent and luck, they may end up with their own source materials resulting from their own crosses. These may then become the source for many of their new entrants. This is seen in Moldovan, Davidson, Petit, Stamile, Apps and others.

3. Promote yourself to the extreme if you want awards. I have often told those seeking business advice that "to get on the bus you must be standing on the corner, it just does not drive into your bedroom.." Thus for those who seek glory, they must get into the market and promote themselves. Looking at Stamile one sees a great promoter, and in turn one who has obtained many awards. The awards track is a club, and as a club one must work their way up to the top. That does not mean in any way that those who hybridize for the sake of hybridizing are to be marginalized. In many ways they are like gold nuggets, they can be mined for new product.

4. Create goals but be pragmatic and opportunistic. One can set out seeking doubles and find spiders. Thus having rigid goals will not necessarily result in a good outcome.

5. Look at the fringe versus the center. Decide where to play. The fringe is where the new introductions are, they are at the point of introducing the new gimmick, a metallic edge, a speckled eye, and many of the forms as described by Peck. In contrast there is the player in the center who is looking for good horticultural product. This means a good and hardy grower, a good and consistent display plant, and one which can be combined with others to create a palette. Again I think of Mahieu as a player in this field.

These are not rules from anyone specific but they are a condensation of what has been heard from many hybridizers. One need look no farther than Apps to see a superb middle of the road hybridizer, or Stevens, while a generation ago created a set of plants which had more than stood the test of time.

8 History of Hybridizing

Before proceeding it is useful to provide some insight as to the progress of hybridizing in Hemerocallis over the past hundred years or so. We will rely upon both secondary and primary sources in presenting this history. For example, with Stout we have both his writings and the anecdotes from those at the New York Botanical Garden where he performed his work. With many of the contemporary hybridizers we have had first hand conversations. One thing seems common; they all have an intuitive feel for mixing the plants to achieve their intended goals, which most often in innovation of form.

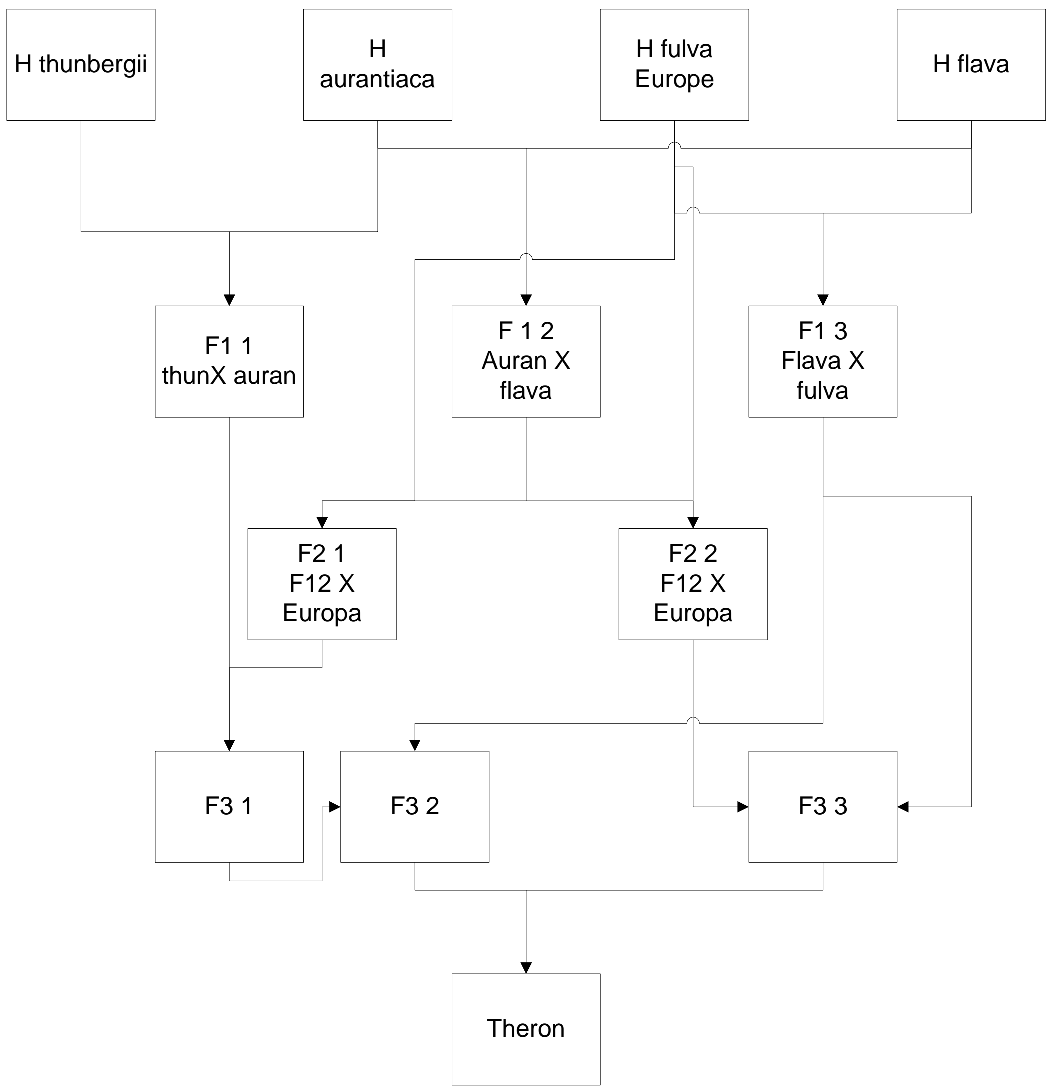
8.1 Early Hybridizing Developments

In the early years, from the late 19th century onwards towards the mid 20th century we rely primarily upon Munson and Stout.

8.1.1 Stout

Arlow Stout performed his research at the New York Botanical Garden in the borough of the Bronx at the northern end of the City of New York. The Garden lays aside the Bronx River, which flows south and at the point where the Garden lies it bisects the Garden and the Bronx Zoo. This piece of land is the only preserved land in the City of New York having never been clear cut. Stout worked there in the first half of the twentieth century when this was still a somewhat rural part of the City. He had adequate land to grow his many hybrids. He communicated with many who went on trips to the Orient collecting plants and he was thus able to obtain and propagate an enormous variety of the genus. He published his book on Daylilies in 1934.

One of his classics is Theron which he shows in his book as a cross as follows:



The following is an example of a few of his early hybrids. He worked tens of thousands of crosses, learning in detail what would work with what cross, and diligently recorded all of his cross data. It was a masterful effort in science.

Stout

Mikado 1929

Autumn Minaret

Theron

Rajah 1935

Buckeye 1941

Dauntless

Elfin

We can see in the above hybrids some of the features that were to come. The red of Theron was a first in a deep red color with a sense of pureness to the color as compared to a mottled H aurantiaca. The eye zone of Autumn Minaret was the beginning of the eye zones we see again and again. Dauntless also has such an eye zone. Buckeye and Mikado have stronger eyezones.

8.1.2 Others

During this early period there were many more amateur hybridizers. The source materials were few and the communications between the hybridizers was limited and slow. It was also a period of the Depression and the Second World War.

Munson details some of the early hybridizers. He speaks of Stout, Yeld, Wheeler, Taylor, Nesmith, Connell, Lester and Milliken. We show some of this early work starting with Mead and Hyperion, still a standard. This is shown below.

Mead

Hyperion 1924

Nesmith, Elizabeth Nesmith, who Munson calls Miss Betty, introduced Potentate in 1943. The is shown below. It was one of the first deep red flowers, and in many ways is a departure from many of the others bred until that time. It has a clarity and form which sets it apart and begins a road towards a collection of reds and purple. Munson calls it a violet-plum, and indeed it can be seen that way. However we have used it as a base for reds as well.

Nesmith

Potentate
1943

The second hybridizer is Bechtold and he introduce Kindly Light one of the earliest spiders and a plant which sees continuing use as a source for spider forms.

Bechtold

Kindly Light 1949

Munson mentions Kraus, but in the Middle period and we place him in the early one for several of his introductions. Below we show Yellowstone, another plant which is still collected and grown extensively.

Kraus

Yellowstone 1950

8.2 Middle Ages of Hybridizing

The Middle Ages for hybridizing was from 1950 to 1975 for Munson. We have expanded this until 1980. In his discussion Munson includes Kraus, Hall, Claar, MacMillan, Spalding, Childs, and for the Tets, Peck, Marsh, Fay, Reckamp, Moldovan, Munson. We will look at the work of a few others during this period. Specifically:

- Peck
- Winniford
- Stevens
- Davidson

8.2.1 Childs

Childs introduced three interesting flowers. Catherine Woodbury is a true pastel and still is an attractive addition to any garden. It is also a source for hybridizing diploids with pastel structure. Ice Carnival is a white which has been used by many others for attaining a purer white as well as a base for blending other colors. Genetically the question is how one creates a white, possibly by just turning off all anthocyanin pathways or by adding the correct balance.

Childs

Catherine Woodbury 1967

Ice Carnival 1967

Try It 1972

8.2.2 Hall

Hall was a mid-west hybridizer who sold his stock ultimately to Wild. His early hybrids as shown below demonstrate the initiation of bi-colors as well as blending. The bi-color he developed has been used as stock in many future bi-colors. The blend is an attractive base in some limited hybridizing.

Hall

Magic Dawn
1955

Precious One
1967

Orchid Pink 1955

8.2.3 Marsh

James Marsh worked in both Dips and Tets. One of his most significant contribution was Prairie Blue Eyes, one of the earliest attempts to achieve a blue color in daylilies. Also Prairies Moonlight is a very light yellow verging on white. These two were done in the period of 1965-1970. He then started his hybridizing in Tets with the Chicago series. There he achieved a great deal of success with the reds and with pastes, such as Chicago Catelleyea.

We show several of his introductions below. They all possess good solid growth characteristics and present very well in almost any garden.

Marsh

2N Diploids

4N Tets

Prairie Blue Eyes
1970

Prairie Moonlight
1965

Chicago Fire
1972

Chicago Atlas
1975

Chicago Brave
1976

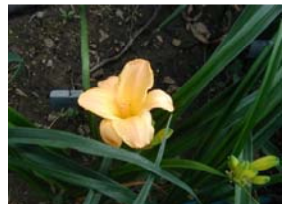
Chicago Catelleyea
1980

Marsh shows great diversity in color as well as form in this period. The Prairie series were all Dips and the Chicago all Tets. The difference in added sophistication with the Tets is obvious as you look at them side by side. However the simple and direct clarity of the Dips keeps them in circulation and for Dip hybridizers they are a base for continuing the subtle elements that Marsh introduced. The Prairie Blue Eyes has been used extensively for the introduction of Impressionistic color combinations.

8.2.4 Peck

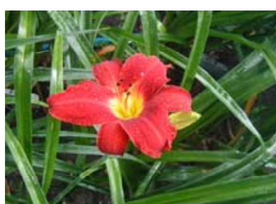
Virginia Peck, as states Munson, is a breeder from Tennessee. She has worked with Tets for many years and during this period made many important introductions whose use in hybridizing is still used. We show several of them below. Wine Bold is a rich dark red flower with good growth and it provides the basis for many dark red hybrids. June Wine is an eyezone which is also the basis for many eyezone plants. Jog On and Scarlett Kettle are rich bright reds which also can be used to infuse color into plants.

Peck


Etched in Gold 1972

Wine Bold 1972

Tammas 1972


Scarlett Kettle 1976

June Wine 1976

Jog On 1976

From 1972 through 1976 the reds introduce by Peck were the basis for reds used by many other hybridizers as well. One can see in the above the less than subtle difference in the four reds she introduced during that period.

8.2.5 WINNIFORD

Ury Winniford of Dallas Texas introduced 205 hybrids from 1968 thru 1990. Two of his early introductions are shown below. They are the tinted eyezone Tixie which is small but a good growing plant even in the north and Brutus which has a unique cup like form and is an aggressive grower. Winniford in this mid period introduced many hybrids and they have interesting forms and shapes.

Winniford

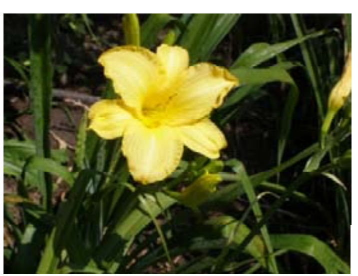
Brutus 1975

Tixie 1974

8.3 WILD

Wild purchased the Hall crosses and added to them Some examples of his introductions are shown below:

Wild



Alice in Wonderland

American Revolution

Border Beauty

Coming Your Way

Dawnbreaker

Ginger Whip

The above hybrids were developed for mass commercial sale and were sold to a mass market ate generally low prices. However they do have reasonably attractive quality given the time of introduction and do find homes in many gardens.

8.4 Recent Hybridizers

To understand the way modern hybridizing is accomplished it is useful to have a better understanding of the hybridizer's techniques and goals. From a 1957 article speaking to the evaluation of the daylily the authors recounts the considerations that Stout applied to the selection of hybrids. He specified them as:

1. The plants should have winter hardiness.

2. The plant should bloom for a long season.

3. Flower color should not bleach out and petals and sepals should not curl or wilt prematurely.

4. Flowers must drop quickly after bloom on their own.

5. Flowers should stay open in the evenings.

6. Flowers must sit high enough above the foliage so as to be seen.

7. Scapes should be neither too heavy to overwhelm the plant or too thin to allow drooping.

8. Foliage must be full, lush and green.

These requirements say much about the plant as a whole and little about the flower in particular. The tracking of new hybrids of the plant can be accomplished via the AHS award process. There are several steps in that process.

Step 1, Junior Citation: This is awarded to a plant which has not been registered for more than a year and is frequently even awarded to an unregistered cultivar. This is a regional awarding process and it attempts to reward the newer introductions.

Step 2, Honorable Mention: This award is the next step in cultivar evaluation and now moves from possibly just one local region to a minimum of four or more regions. A cultivar must receive fifteen or more votes from Judges to receive this award. To be eligible the cultivar must have been registered for at least three years.

Step 3, Award of Merit: According to AHS this is awarded not only for a cultivar's distinction and beauty but also for its ability to perform well over a large geographical area. Twelve awards are made each year. To be eligible a cultivar must have received an Honorable Mention for three previous years. For example in 2007 there were the full twelve Awards of Merit.

Step 4, Stout Silver Medal: The award is given annually to a cultivar which must have received at least two prior Awards of Merit. The Stout Medal is the highest award from the Society. The list of past winners is shown in the Table below.

2007 LAVENDER BLUE BABY (Carpenter, 1996)
2006 ED BROWN (Salter, 1994)
2005 FOOLED ME (Reilly-Hein 1990)
2004 MOONLIT MASQUARADE (Salter, 1992)
2003 PRIMAL SCREAM (Hanson, C. 1994)
2002 BILL NORRIS (Kirchhoff, D. 1993)
2001 IDA'S MAGIC (Munson, I. 1988)
2000 ELIZABETH SALTER (Salter 1990)
1999 CUSTARD CANDY (Stamile 1989)
1998 STRAWBERRY CANDY (Stamile 1989)
1997 ALWAYS AFTERNOON (Morss 1987)
1996 WEDDING BAND (Stamile 1987)
1995 NEAL BERREY (Sikes 1985)

1994 JANICE BROWN (Brown 1986)
1993 SILOAM DOUBLE CLASSIC (Henry 1985)
1992 BARBARA MITCHELL (Pierce 1984)
1991 BETTY WOODS (Kirchhoff 1980)
1990 FAIRY TALE PINK (Pierce 1980)
1989 BROCADED GOWN (Millikan 1979)
1988 MARTHA ADAMS (Spalding 1979)
1987 BECKY LYNN (Guidry 1977)
1986 JANET GAYLE (Guidry 1976)
1985 STELLA DE ORO (Jablonski 1975)
1984 MY BELLE (Durio 1973)
1983 SABIE (MacMillan 1974)
1982 RUFFLED APRICOT (Baker 1972)
1981 ED MURRAY (Grovatt 1971)
1980 BERTIE FERRIS (Winniford 1969)
1979 MOMENT OF TRUTH (MacMillan 1968)
1978 MARY TODD (Fay 1967)
1977 GREEN GLITTER (Harrison 1964)
1976 GREEN FLUTTER (Williamson 1964)
1975 CLARENCE SIMON (MacMillan 1966)
1974 WINNING WAYS (Wild 1963)
1973 LAVENDER FLIGHT (Spalding 1963)
1972 HORTENSIA (Branch 1964)
1971 RENEE (Dill 1962)
1970 AVA MICHELLE (Flory 1960)
1969 MAY HALL (Hall 1957)
1968 SATIN GLASS (Fay 1960)
1967 FULL REWARD (McVicker 1957)
1966 CARTWHEELS (Fay 1956)
1965 LUXURY LACE (Spalding 1959)
1964 FRANCES FAY (Fay 1957)
1963 MULTNOMAH (Kraus 1954)
1962 BESS ROSS (Claar 1951)
1961 PLAYBOY (Wheeler 1954)
1960 FAIRY WINGS (Lester 1952)
1959 SALMON SHEEN (Taylor 1951)
1958 HIGH NOON (Milliken 1948)
1957 RUFFLED PINAFORE (Milliken 1948)
1956 NARANJA (Wheeler 1947)
1955 PRIMA DONNA (Taylor 1946)
1954 DAUNTLESS (Stout 1935)
1953 REVOLUTE (Sass 1944)
1952 POTENTATE (Nesmith 1943)
1951 PAINTED LADY (Russell 1942)
1950 HESPERUS (Sass 1940)

From this list it is clear those hybridizers such as:

8.4.1 Stamile, Patrick and Grace

Patrick Stamile has 5 Stout Medals, 27 Awards of Merit and 115 Honorable Mentions. He is a prodigious hybridizer who started his introductions in 1984. He initially started his hybridizing in 1977. Patrick Stamile initially started his growing on Long Island and in 1993 he moved with his wife Grace to Florida. Since then his introductions have a southern bent and in many ways have become southern hybrids. Patrick Stamile represents a standard for hybridizers, namely going out and making contact with those who have achieved recognition and success, seek their advice and technique, and obtain hybridizing materials and then focus on their hybridizing. Grace Stamile has been focusing on hybridizing miniature and blue tinted hybrids for twenty years. In 1989 she obtained her first hybrid called Coming Out Party.

It was the beginning of a blue period. She used several hybrids which had both blue and small flowers to combine them to seek out the traits she was seeking. She has 30 Honorable Mentions. Grace's approach is quite focused using They have been in Enterprise, FL for the last fifteen years. The approach used by both seems to be standard but a standard using their own stock and expertise. They have several watermarked type of flowers and it is clear looking at the parentage that they have achieved good mixing by using the incremental strength of the breeding parentage. One may try to intuit a breeding plan or strategy but it appears to be more a combined mass selection approach yet using pedigree parents. That is choosing the parents and then grows as many seedlings as possible and chooses the best. There does not appear to be any complex backcrossing or the like.

Stamile 1

Vanilla Candy 1990

White Crinoline 1992

Custard Candy 1989

Tigger 1989

Stamile 2

Mystical Rainbow 1996

Strawberry Candy 1989

Wineberry Candy 1990

In the above we show several classes of the Stamile intros. The Vanilla Candy and White Crinoline are two of the whites; Custard Candy is part of his eyed Candy series.

The Mystical Rainbow is the only Stamile introduction in the above which originated from the Florida period. The other Stamile hybrids are from his time on Long Island.

8.4.2 KIRCHHOFF

David Kirchhoff is another Florida hybridizer who in 2006 moved north to Kentucky. He comes from a long line of horticulturalists and growers and has been hybridizing for many years now. He has reds, oranges, dips and Tets. Kirchhoff first crossed a daylily in 1958. Kirchhoff has 107 Honorable Mentions, 17 Awards of Merit and 2 Stout Medals. Betty Woods and Bill Norris are his two Stout Medal winners. His most recent work is on doubles like Barry Goldwater, an orange almost peony like flower which has some reddish edging. It is clear that the attempt here is to take forms which become distinct and enhance them with a different color while keeping the double form[51].

[51] **See http://www.daylilyworld.com/dw-intro--pages/barry_goldwater.htm "Descended from an out cross breeding George Rasmussen's TIGER PARADE to our LAYERS OF GOLD. Ninety nine percent double"**

His stated approach was an outcrossing method with doubles and the outcrossing introduced additional genetic diversity. Kirchhoff has a partner one Mort Morss, who has been hybridizing with Kirchhoff for over thirty years, since 1971. One of his recent introductions is Curtis Montgomery which is a beautiful bicolor with a watermarked eye and ruffled petals. The petals are a reddish orange and the sepals are peach. It appears to be an aggressive grower.

Kirchoff

Bill Norris 1993

The above is an example of Kirchhoff. The classic one is Bill Norris, an award winner. It is a pure deep yellow with full petals and sepals and ruffled edges. Depending on where it is grown it will do well or poorly. In our experience it does well in New Hampshire and poorly in northern New Jersey soils.

8.4.3 Moldovan

Steve Moldovan and his partner and successor Roy Woodhall did their hybridizing in Avon, Ohio, and west of Cleveland and near the lake. It is a cold and snowy environment in the winter but can be somewhat moderated in the summer. It is not Florida in any way of the imagination. Steve Moldovan passed away on July 14, 2006. Roy Woodhall continues the work of Moldovan. He was 68 and he had been hybridizing almost all his life. He held a graduate degree in Horticulture from Ohio State University he introduced many exceptional hybrids. He had 43 Honorable Mentions and 6 Awards of Merit. The key thread that seems to have led Moldovan was his early contact with the hybridizers of the previous generation; Reckamp, Munson, Fay, and many of the now classic hybridizers. This, along with his own training, seems to have given him an exceptional basis for developing his own technique as well as his own line of plants.

The following are four classic Moldovan introductions. They all show a pastel like character and lack the pattern formation he sought at latter times.

Moldovan

French Tudor

Seurat

Strutter's Ball

Tachibana

One of Moldovan's best hybrids, Strutter's Ball, is a cross between his own Houdini and Munson's Damascus Velvet. All three are reds and all three have a green gold throat. Strutter's Ball is an exceptional bloomer and is well branched with many buds. It had become a key element in many of the Moldovan crosses.

In fact as Woodhall has said of the techniques he has developed working with Moldovan the one which is often the most important is to generate one's own parent hybrids, those with characteristics that make your showings different and use that source a

Moldovan was one of the first in the area of Tets and also was one who worked with the many pastels we have come to see out of the crossings, again and again. Recently one can see in his final hybrids the introduction of some bicolors and some of the shapes and coloring common in many of the other commercial hybridizers.

In the article by Fitzpatrick on Moldovan just before his death she recounts the rules he promulgated for hybridizers[52]:

1. Plant many seeds but be prepared for the retention of very few, one out of a thousand.

2. Outcross to hardy cultivars to ensure that the perennial does not become an annual.

3. The results of a cross are never certain, and in fact never imagined.

4. Always be aware for special little traits. They can be used again and again and introduced into new crosses.

5. Plant seedlings in the ground. Let Nature do its pruning.

Moldovan's rules are to be well taken. The hybridizer seeking a truly sustainable set of greatly appreciated hybrids will take them to heart. We expand on Moldovan's five rules below:

[52] See Sharon Fitzpatrick, Steve Moldovan's Quest, The Daylily Journal, Fall 2005, pp. 312-323.

Moldovan Rule	*Implication*
Plant many seeds	This is the rule that says you increase your chances with larger numbers to select from. You will look only for one in a large number. You may see one in a hundred as something to consider and one in a thousand to keep.
Outcross with Hardy plants	Outcrossing, the crossing with stronger and dramatically different hybrids, and some would say even species, puts genetic diversity back to the plant. Excessive inbreeding will enforce certain characteristics but will also most likely enforce weaknesses that will be highly negative for the plant. Outcrossing, however, will also result in getting the dominant genes back in the pool, and that return of the dominant may wipe out the characteristic we had been seeking. However, we know the gene we wanted to keep may not appear in F1, it will, if it survived appear in F2. This when outcrossing, remember to continue to F2 in all cases.
Crosses are Never Predictable	Despite what we try to say regarding the genetics of plants, the statements hold only in the large, namely on average, and when looking at the hybridizing results we all too frequently select the outliers. The outliers are those with the special traits. Then we try to build on them, not on the traits of the average.
Look for Special Little Traits	Look at each and every resulting cross.
Let Nature prune.	This is an extremely important rule for northern hybrids. For, example, it is well known that many southern hybrids will die off when taken too far north. Whereas if one takes a northern plant and crosses it and lets it be selected for survival in the winder, true hardening off, then what results is a plant stock with increased hardiness.

There is a sixth Moldovan rule, one which he based his early days on; have acquaintances that are highly respected and learn from them, use their stock to start and build on their work. For Moldovan it seems it was Reckamp, Munson and Fay. Between the three there were 226 Honorable Mentions, 33 Achievement Awards and 4 Stout Medals. Those three were superb mentors, and mentoring in the field seems to be a major driver.

8.4.4 MATZEK

Matzek is a New Hampshire hybridizer who has made certain introductions which contain the more complex patterning. Several of these are shown below. These are the Windham series and are a quite attractive set of eyed and patterned flowers with edges. The Windham comes in several colors and we depict three in those below.

Matzek

Windham Caress

Windham Masquerade

Windham Orange

8.4.5 Apps

Darrell Apps has all of his degrees including a PhD from University of Wisconsin. He has finally retired from Woodside Nursery in Bridgeton, NJ after decades as an active grower. Apps also has journeyed to the far reaches of Asia in search of the Hemerocallis species, unlike many of the other hybridizers, who have moved from species into the complex and hectic world of multigenerational hybrids. He has introduced hundreds of hybrids and his first was Nittany Mountain Summer in 1975.

FIGURE 91 APPS NITTANY MOUNTAIN SUMMER

The above shows Nittany Mountain Summer as a simple red with a gold throat. He has won 30 Honorable Mentions, 2 Awards of Merit. Apps has a breeding strategy which looks at the total plant, and this includes leaves, scape, branching, and bud count. The plants he has hybridized are extraordinary in a Stout like manner; they are not just pretty pictures, looking solely at the flower but complete structures.

Dr. Darrel Apps is clearly one of the foremost hybridizers over the past forty years. Until 2007 he also was a grower of massive amounts of daylilies until his retirement. His work is an example of a broadly based hybridizer who sought to develop many of the fundamental elements of the genus in all his introductions. He developed hybrids which had good form, structure, color, bloom strength, and he did not focus especially on the bizarre and strange forms. He had a few doubles, few spiders and generally tried to avoid the fads. The following is a chronological list of some of the hybrids we have grown.

Hybrid Name	*Ploidy*	*Intro Date*
Nittany Mountain Summer	2N	1975
Nouveau Riche	2N	1990
Doll Maker	2N	1992
Ebony & Ivory	2N	1992
ORNATE RUFFLES	2N	1992
Royal Frosting	2N	1993
Confectioners Delight	2N	1995
Justin George	2N	1995
Bridgeton Born	4N	1997
Dazzling Discus	2N	1999
Double Intrigue	2N	1999
Better Rum	4N	2000
In the Flesh	2N	2000
Bridgeton Finesse	4N	2001
Luminous Bouquet	2N	2001
Woodside Common	2N	2001
Eager Beaver	2N	2002
Bridgeton Hoopla	4N	2003
Just the Two of Us	2N	2005

The following Figures depict several of these in alphabetical order. What can be noticed in the development are that early on such flowers as Nouveau Riche and Doll Maker are almost mono-color but have tremendous blooms, strong scapes, many buds and good branching. What Apps seems to be focusing on was good underlying form and structure.

In the latter stages with Bridgeton Hoopla and Bridgeton Finesse we see the use of eyezones and with edging on the flowers. However the underlying strength of structure ensures the new form is well supported.

One can see the progression from the Nittany Mountain Summer simplicity to the Bridgeton Hoopla complexity the change not only in his breeding style but in what the market is demanding. There is the growth of ruffles and ridges, the eyezones with the watermarks, the less than subtle colors. Notwithstanding the complexity, however, each Apps introduction also has significant branching and bud count. That quality is a sine qua non of his introductions.

FIGURE 92 APPS PLANTS NO. 1

The second group of hybrids are shown below. These are some with the simplicity of his early introductions, simple color but elegant form and exceptional growth characteristics.

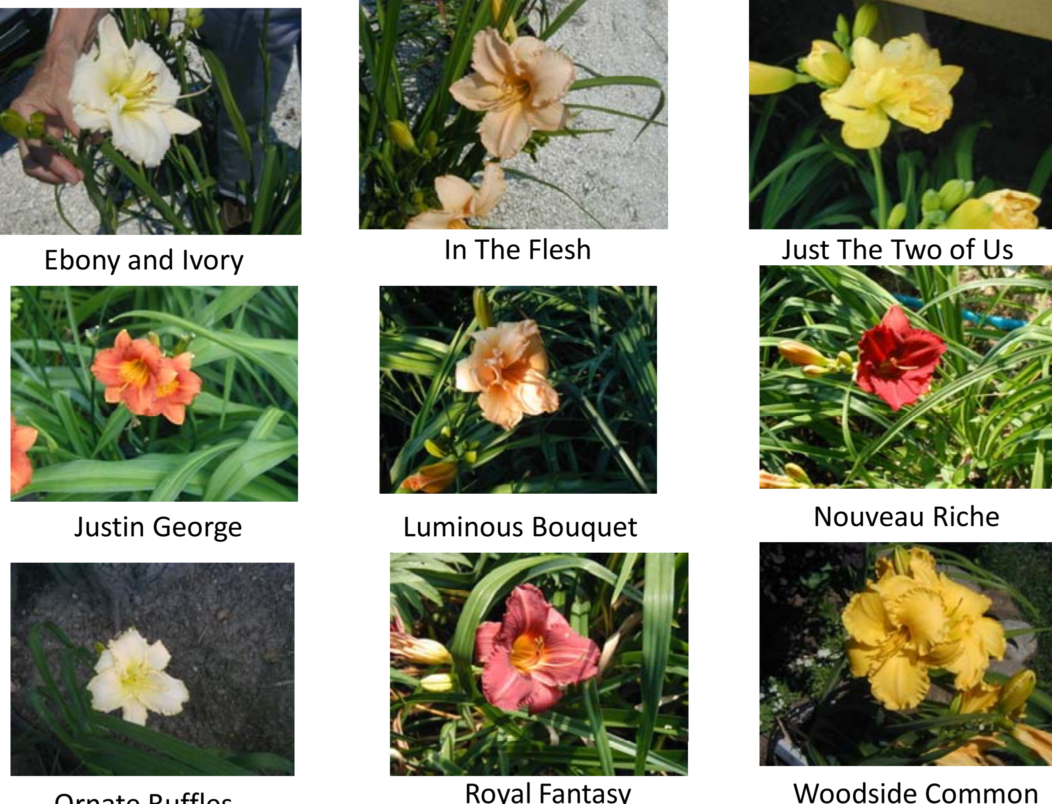


FIGURE 93 APPS PLANTS NO 2

Apps hybrids have certain enduring characteristics. They are:

Excellent form: The plants have well branched scapes with many buds per scape. The scape is strong while not overpowering. It provides an excellent base for presenting the flower. Apps seems to have been very consistent in developing hybrids which sustain that virtue.

Color Intensity: His flowers all have a clarity and intensity that make them stand out, not because of complexity but due to the clarity. Woodside Common is a rich gold yellow and it is the strength of that richness that makes it sit and be noticed.

Growability: The plants generally grow very well. They lack the fragility of the southern hybrids and contain durability to the northern winters. They grow and replicate vegetatively each year in a very productive manner. Unlike many of the fancier hybrids, especially those with complex coloration and/or from Florida, the Apps plants seem to have vigorous annual growth thus allowing extensive vegetative propagation. Perhaps

pricing should be related to how well it can be reproduced vegetatively and not how fragile it as a grower.

8.4.6 Stevens

Don Stevens was from southern New Hampshire and he befriended Bob Seawright who had a growing area in Carlisle. MA. It was from Bob that I received my first batch of daylilies. It was also from Bon that I have many Don Stevens hybrids. Stevens was born in 1930 in New Hampshire and taught in the Bedford, MA High School. Bedford adjoins Carlisle on one side and Lexington MA on the other. Don's hybrids encompassed a wide variety of form, color and shape.

One of the more famous of Stevens's hybrids is the very late blooming Sandra Elizabeth, which in northern New Jersey blooms in early September. It is very healthy and strongly scaped plant with a yellow flower with extreme clarity. It just fills the garden after all of the others have gone their way.

FIGURE 94 SANDRA ELIZABETH

Don Stevens worked along-side Bob Seawright of Carlisle Mass. In fact they jointly hybridized several plants. The Stevens plants are quite sophisticated and are all strong growers and have good bud counts and a balanced color subtlety as well.

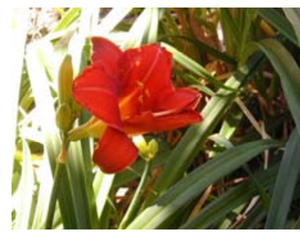
In many ways the Stevens introductions during this period are middle of the road benchmarks. Super Child is an aggressively tall Tet with a very thick scape and tall and large flowers. It almost speaks Tet in its presentation. The following Figure depicts the many introductions by Stevens in the 1970s.

Royal Kingdom and Outrageous show the growing interest in eyezones. The breath of the Stevens introductions is quite wide and they are generally good Northern flowers.

Stevens

Holiday Delight 1978

Outrageous 1978


Fire Tree 1979

Super Child 1979

Royal Kingdom 1980

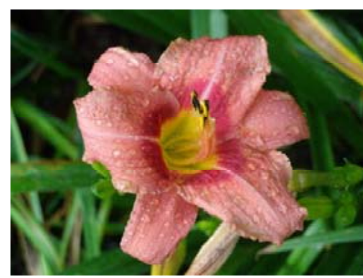
Something Royal 1980

The above are several of the Stevens introductions. One should remember he did these in the 1970s and in addition he only hybridized over an eight year period. The results are amazing for the time and the period. Super Child is a classic standout where Steven created a strong scaped Tet and a blossom that at the end of the season truly stands out.

Some additional Stevens's plants are shown below. Love Festival and Juniper Chase are superb sources to hybridize on because the plant has strong scapes, many branches and many buds. The colors are strong and can be used with some of the more recent introductions.

Stevens 2

Juniper Chase



Lilting Lady

Love Festival

Rachael Hope

Yes

Royal Kingdom presaged many of the eyezone plants of the 1980s and thru the 1990s and is used as parentage in many of these lines. Outrageous also is a deep eyezoned red flower and although not as big as Outrageous has great presence.

8.4.7 DAVIDSON

Clyde Davidson of Decatur Georgia hybridized from 1962 through 1995. His classic is Decatur Apricot, a strong aggressively growing peach or apricot colored Tet. He had registered 184 hybrids and the variation can be seen in a few shown below from his earlier period, Decatur Cherry Smash is a red wine colored Tet with a dark deep red eyezone. It is recurved and presents very well in the garden. It is not as strong a grower as is Decatur Apricot but does well.

Davidson

Decatur Apricot 1977

Decatur Cherry Smash 1980

Decatur Dictator 1979

FIGURE 95 DAVIDSON DECATUR SERIES

The Davidson Decatur series as shown above are also a series in the 1970s and they are a strong set of good growing Tets. Decatur Apricot has been used as a parent for many Tet lines and it has the dark peach, apricot, color and strong branching and bud count.

8.4.8 Petit

Ted Petit is known, along with his partner John Peat, as the authors of a well organized and successful book on the general areas of the daylily. To a great degree Petit is a "leading edge" hybridizer whose success seems to come from noticing the small changes and nuances and building upon them, using breeding techniques which drive the subtle effect deeper into his breeding line. He states that Munson was an influence on him and that especially the comment by Munson where he desired to have an award named for him for the best patterned plant[53]. He continues he recounting of his conversations with Munson by stating that Munson felt the future of hybridizing was in patterns, for other characteristics such as ruffles would just drive the plant to the extreme. Patterns were where the new elements of near endless creativity could be attained.

These trends in patterning are then shown in some detail by Petit in both his work and that of others. He classes the patterns as follows:

[53] See Petit, Daylily Journal, Summer 2007, pp. 125-141.

Appliqué Throats: This is what Petit calls a pearl like patina in the throat. He attributes some of these to Munson. The pattern appears as an application on top of the flower and not coming from within.

Mascara Eyes or Bands: This is the eyezone which has a darkening or contrasting color on the interface region. Again this was a Munson construct. Early versions of this patterning are by Salter. In many ways these flowers appear as if one had dropped food coloring water on a cloth and the eyezone diffuses outwards. There is lack of true clarity. In view of the Turing model for color these flowers and this patterning provide excellent example of true diffusion.

Inward Streaks: This is inward veining especially in the eyezone portion.

Concentric Circles or Bands: This is the alternate to the Inward Streaks by having circular bands.

Washed Eyezones: These are the "running" out of the eyezone in an almost random but limited fashion.

Stippling: This is a dotting effect, which Petit also calls speckled. The coloration appears as if it were done in some impressionistic painting. The colors are not blended but are interspersed.

Metallic Eyes: Like the Appliqué Throats the Metallic Eyes appear as if they have metal specks residing on the top of the eye pattern.

Veining: These have highly contrasted vein patterns.

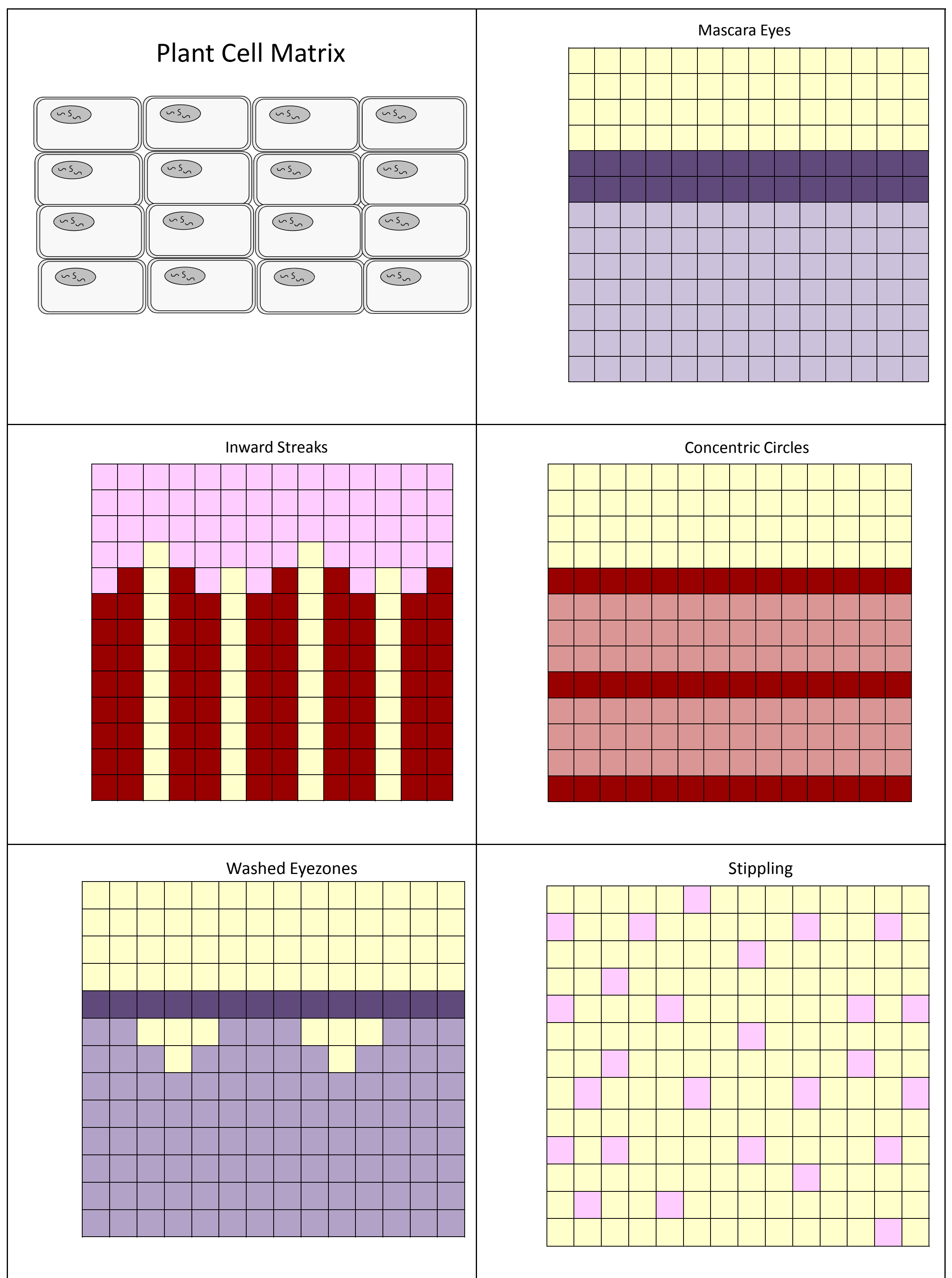
Rainbow Edges and Midribs: These have edged and midribs where the color variation is a complex set of different colors. This presents a very important model to apply the Turing approach to. It may allow for the inversion problem to seek a solution, for it shows how the instability of the secondary pathways can be controlled.

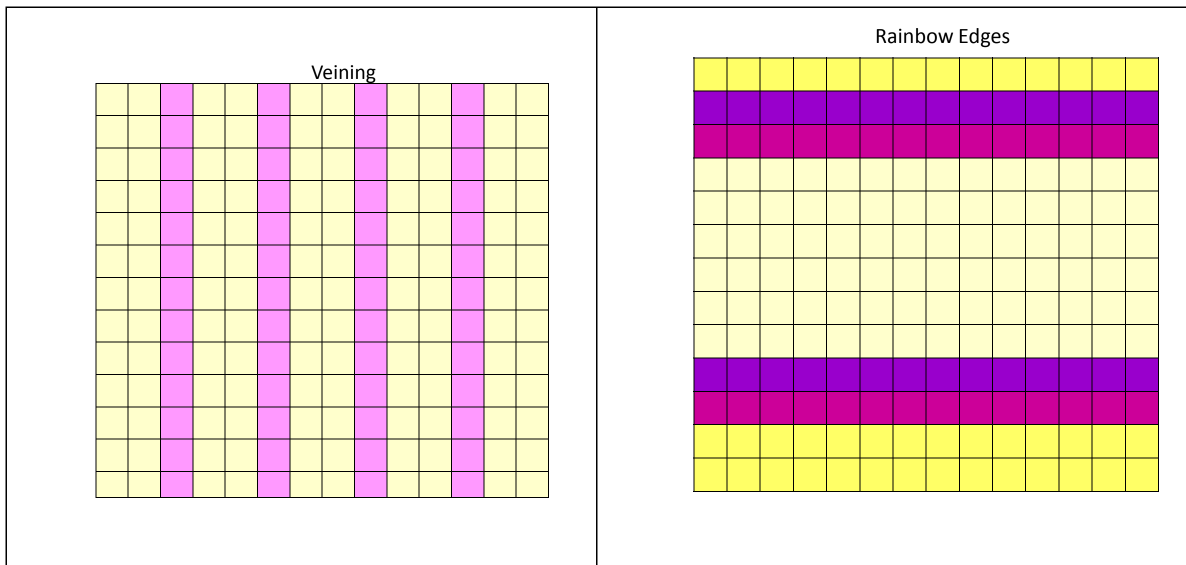
Narrow Formed: These are the contradistinctions of the round daylily. Here form rather than color become a variant.

Others: Petit also presents a collection of yet to be classifies forms.

Characteristic	***Turing Model***
Appliqué Throats	Unknown mechanism
Mascara Eyes or Bands	Demonstrates multiple layers of low spatial frequency outward growth of color.
Inward Streaks	If flower grows outward then the flow of control is unstable across new rows of growth.
Concentric Circles or Bands	If flower grows outward then the flow of control is unstable between new rows of growth.
Washed Eyezones	Ultra High intercellular instability, with almost localized oscillations allowing high spatial frequency of color change.
Stippling	High intercellular instability, with almost localized oscillations allowing high spatial frequency of color change.
Metallic Eyes	Unknown mechanism
Veining	Demonstrates multiple layers of low spatial frequency lateral growth of color.
Rainbow Edges and Midribs	
Narrow Formed	Not Applicable
Others	Not Applicable

Petit uses the sources of this innovative color patterns in his hybridizing as does his partner Peat. These color schemes provide a unique basis for the validation of the Turning model. We show the abstractions of these patterns reflected on a cellular matrix as follows.


Plant Cell Matrix
Mascara Eyes
Inward Streaks
Concentric Circles
Washed Eyezones
Stippling

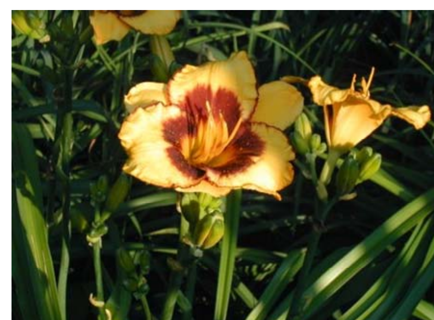


One can note that each of these becomes a Turing model with certain points of instability in a periodic manner. One can predict that there could be an almost unlimited number of such patterns depending on the inbreeding of the gene combinations controlling the stability points.

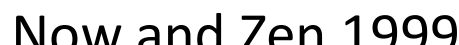
8.4.9 Hanson

Hanson has 1 Stout Silver Medal, 4 Awards of Merit and 29 Honorable Mentions. His Primal Scream is the one for which he received the Stout Medal. In the figure below we show two others. One is Now and Zen, an eyed and edged plant which grows modestly up north and Sea Hunt which is a watermarked purple tinted flower.

Hanson



Sea Hunt 1999



Now and Zen 1999

Primal Scream 1990

8.4.10 Mahieu

Although not an award winner as yet, the plants by Mahieu have an interesting turn. Mahieu is an artist and he brings an eye for subtle color to his introductions as well as an exciting form. Furthermore Mahieu is attracted to the species, especially H citrina and H altissima. He has focused on what he calls the "architecture" of the plant, and in that context he is building on the Stout hybrid Autumn Minaret, which stands tall and quite distinctively in any garden at the end of a season. He wants to emphasize in his breeding the entire plant, and to do so has brought to his crosses the character and strength of not only citrina and altissima but H hakuunensis and H dumortieri.

Mahieu states that he seeks to "put huge blooms …with heavy texture on tall scapes…". Indeed, that is what he has accomplished. Unlike the main stream hybridizers like Munson, Petit, Stamile, and others, Mahieu represents a branch of hybridizing which seeks the new and innovative by drawing back upon the much strength of the original species. Mahieu is an artist and one can see his pallet in his crosses. They are simple, yet elegant, colorful, yet not extreme, and they catch your eye as you enter. They have the subtlety of the impressionists while having the stature of the species. The species is always not very far behind what he has presented.

Mahieu is an example of the hybridizer who brings back those dominant and nature preserving genes which have been driven out by Petit and the others who are seeking the in extremis flower. It is not that either is better or worse, Judges decide what is currently in vogue, yet they both show the versatility of the genus.

8.4.11 Joiner

joiner has developed many doubles which we show four of them below. They generally can be used to set seed and can result in doubles in their crosses. The ones we have are the lighter one since they generally are amenable to crossing with colors.

Joiner

Francis Joiner 1988

Tall and Proud 1994

Madge Cayse 1991

Jean Swan 1993

8.4.12 McGarty

The following are a few of the hybrids introduced by the author. The author has focused in developing strong northern hybrids with eyezones and pastels. The following are examples of such.

Maja's Tinkerbell is an attractive and quickly multiplying blend with a large flower with reddish tint. It has the habit of standing out over all the other flowers for a six week period. Florham Peaches and Cream is more likely a great horticultural flower, with rapid expansion, a peach color with very soft undertones and vey sustainable in the garden.

McGarty 1

Maja's Tinkerbell

Florham Peaches and Cream

Kris' Kindness

Mr Brown

The following are examples of eyezones we have hybridized. Rita's Sunrise was created in 1994 and has demonstrated a fast growth pattern. It divides rapidly and is a tall and impressive flower. It is between a horticultural plant and an exhibit display plant. Princess Martina is also a very attractive recent introduction with a strong eyezone and recurved sepals and petals. It has both color and form and this combination makes for an attractive display flower.

McGarty 2

Princess Martina

Rita's Sunrise

Sara's Wink

Happy Hilda

9 CONCLUSIONS

In this book we have addressed the issues related to the genetic control and influence on the coloration of flowers and we have used the Genus Hemerocallis as a vehicle to do so. Rather than this being a book on Hemerocallis and its variants alone or a book on genetic analysis of coloration or even secondary pathways, we have delivered an admixture of the two. In reality one often learns about general principles via studies of specific examples, especially one where we can see and measure the variations in a readily accessible manner. Thus using Hemerocallis was a means to an end, the end being an understanding of genetic control and management.

Also one of the things we find critical is our ability to measure the results obtained, to quantify them as best as we can. Thus the discussions we have pursued on such items as color measurements and gene expression measurement are critical in taking us beyond the Mendellian world of the gene abstraction, and into the world of gene control.

9.1 KEY OBSERVATIONS

There are many observations of a general nature we can make. They can be done with regard to the issues of the genus and with regard to the issues of the procedures, methods and processes we have presented.

9.1.1 GENUS

The overriding question is what are the species and how do we define them. The list of 30, 26, 24, 12 or whatever, are in many ways arbitrary. If we recall that in plants a species line is not so clearly drawn, and that geographical isolation is not even a limit, then the major issue we should resolve is that of what is the species. The recent work of Niklas in Evolutionary Biology is an exceptionally clear statement of this very issue[54]. Species in the plant world are less clearly defined. They are not separated by the inability to reproduce as Mayr had stipulated. The flow of genes back and forth creates more of a continuum of form and function. Clearly the early flowering species may be arguably separate from the late flowering due to temporal separation. However the spatial separation does not establish such a boundary.

The evolution of the various hybrids also presents and interesting focus. The schools of hybridizers seem to be currently dominated by those of the bizarre. They are like autos in the 1950s with ever so larger a fin, a grill, and attachment inside, they lack form and function and go for the extreme. The argument of Munson that patterns would be the

[54] See Niklas pp 63-108.

next focus was spot on except that the patterns are moving to the edge. Yet it is in these patterns that we can hopefully better understand through Turing type models the mechanism of gene control.

9.1.2 Methods and Procedures and Processes

The approach we have taken herein to describe and develop methods, procedures and processes is an engineering approach. Namely we have looked at the genetic detail of the cell and the secondary pathway mechanism and we have abstracted from these the details we believe are adequate to both explain and in turn control or modify the results. This is akin to designing a transistor circuit. One abstracts the quantum physics of holes and electrons in germanium and silicon to see just the input and output of the transistor. The details of the quantum electrodynamics are left as an exercise of the physicist.

Thus it is possible that the molecular geneticist may see we have neglected and abstracted to a degree which would make them uncomfortable. The true question then is have we abstracted too far and in so do have we lost the effective elements of the processes being described. That can only be determined by continual experimentation.

In addition the field of molecular genetics is changing at least by the hour if not by the minute. What we think we know today we may have to revise on the morrow.

We have presented several methods for both analyzing the issues associated with flower color and also for controlling flower color and patterns. We have laid out in Chapter 6 the models for patterning and for color determination; the inter-cell and intra-cell problem. There were assumptions made concerning certain bulk parameters that must be validated. It is unlikely that the model must be changed totally but modifications to incorporate secondary effects are anticipated.

9.2 Unanswered Issues

There are still many unanswered questions in this area. We present several of them here. Although we have tried to present what appears to be a fully connected discussion of the genus and the resulting control over flower color, we have made certain assumptions that can only be validated by extensive experimentation and testing. As with the above issues they fall into those relating to species and those relating to the genetic engineering and analysis tools we have developed.

9.2.1 Genus

Some of the ones relating to the genus issues are:

What is the metric for determining one species from another?

Are there species which are themselves hybrids?

How do we define species in this genus? What is the dividing line?

What is the genetic history of the genus? Can we determine the genetic ancestors from the genetic pool currently available?

9.2.2 Methods and Procedures and Processes

Are the models for patterning robust enough to be predictive of all patterns?

Can patterns yet to be obtained be defined and if so can they be engineered?

Can we determine the constants that determine the patterning metrics? If so can we reproduce patterning by analysis?

Can we synthesize patterns by genetic control of the constants in the Turing model?

Can we engineer flowers for specific colors by reverse engineering the color model? Are there colors which are not achievable given the underlying genetic makeup of the secondary pathways? Namely can we engineer a blue daylily from what is genetically available in the current genus?

9.3 Extensions

The worked covered herein used the Genus Hemerocallis as a vehicle, as a means to an end. Although we used the genus as the vehicle to look at flower color the real issue was that we looked at gene control of secondary pathways. The areas of extension are:

1. Secondary pathway control in many human cancers.
2. Secondary pathway in controlling field crops.

10 References

1. Allard, R., **Plant Breeding**, Wiley (New York) 1960.
2. Atkins, P. **Physical Chemistry**, Freeman (New York) 1990.
3. Baici, A., *Enzyme Kinetics*, The Velocity of Reactions, Biochem Journal, 2006, pp. 1-3.
4. Bartel, B., S. Matsuda, *Seeing Red*, Science, Vol 299, 17 Jan 2003, pp 352-353.
5. Benson D. L., *Unraveling the Turing Bifurcation Using Spatially Varying Diffusion Coefficients*, Jour Math Bio, Vol 37 1998, pp. 381-417.
6. Benson, D. L., et al, *Diffusion Driven Instability in an Inhomogeneous Medium*, Bull Math Bio, Vol 55 1993, PP. 365-384.
7. Berns, R. S., **Principles of Color Technology**, Wiley (New York) 2000.
8. Born, M., E. Wolf, **Principles of Optics**, 4th Ed, Pergamon (New York) 1970.
9. Brown, A., et al, **Plant Population Genetics, Breeding, and Genetic Resources**, Sinauer (Sunderland, MA) 1990.
10. Campbell, A., L. Heyer, **Genomics, Proteomics, and Bioinformatics**, Benjamin Cummings (New York) 2003.
11. Carey, F. A., **Organic Chemistry**, McGraw Hill (New York) 1996.
12. Causton, H. et al, **Microarray Gene Expression and Analysis**, Blackwell (Malden, MA) 2003.
13. Cavalli-Sfroza, L. L., W. F. Bodmer, **The Genetics of Human Populations**, Dover (Mineola, NY) 1999.
14. Chase, M., *Monocot Relationships, an Overview*, Journal of Botany, Vol 91 2004 pp 1645-1655.
15. Chen, T., et al, *Modeling Gene Expression with Differential Equations*, Pacific Symposium on Biocomputing, 1999 pp. 29-40.
16. Chung, M., J. Noguchi, *Geographic spatial correlation of morphological characters of Hemerocallis middendorfii complex*, Ann Bot Fennici Vol 35, 1998, pp. 183-189.
17. Chung, M., *Spatial Structure of three Populations of Hemerocallis hakuunensis*, Bot. Bull. Acad. Sci., 2000, Vol 41, pp. 231-236.
18. Cilla, M. L., D. Jackson, *Plasmodesmata Form and Function*, Current Opinion in Cell Bio, Vol 16 2004 pp. 500-506
19. Crawford, D., **Plant Molecular Systematics**, Wiley (New York) 1990.
20. Cronquist, A., **The Evolution and Classification of Flowering Plants**, New York Botanical Garden Press (Bronx, NY) 1986.
21. Dahlgren, R.M.T., T**he Families of Monocotyledons**, Springer (New York) 1985.
22. Daly, D., et al, Plant Systematics in the Age of Genomics, Plant Physiology, Dec 2001, pp 1328-1333.
23. Dey, P. M., J. B. Harborne, **Plant Biochemistry**, Academic Press (New York) 1997.
24. Dunn, G., B. Everitt, **Mathematical Taxonomy**, Dover (Mineola, NY) 2004.
25. Durbin M. L. et al, Genes *That Determine Flower Color, Molecular Phylogenetics and Evolution*, 2003 pp. 507-518.
26. Durbin, R. et al, **Biological Sequence Analysis**, Cambridge (Cambridge) 1998.
27. Durrett, H., **Color**, Academic Press (New York) 1987.
28. Eisen, M., et al, *Cluster Analysis and Display of Genome Wide Expression Patterns*, Proc Nat Acad Sci 1998 Vol 98 pp 14863-14868.
29. Erhardt, W., **Hemerocalis**, Timber Press (Portland, OR) 1992.
30. Esau, K., **Anatomy of Seed Plants**, Wiley (New York) 1977.
31. Felsenstein, J**., Inferring Phylogenies**, Sinauer (Sunderland, MA) 2004.
32. Fox, M. A., J. K. Whitsell, **Organic Chemistry**, Jones and Bartlett (Boston) 1997.
33. Gitzendanner, M., Soltis, P, *Patterns of Genetic Variation in Rare and Widespread Plant Congeners*, Journal of Botany, Vol 87 2000 pp 783-393.
34. Goodwin, T.W., **Chemistry and Biochemistry of Plant Pigments**, Vols 1 and 2, Academic Press (New York) 1976.
35. Griffiths, A., et al, **Genetic Analysis 5th Ed**, Freeman (New York) 1993.
36. Guerro, C., et al, Analysis of the Expression of Two Thioprotease Genes from Daylily (Hemerocallis spp.) during Flower Senescence, Plant Molecular Biology, Vol 36 1998 pp 565-571.

37. Gusfield, D., **Algorithms on Strings, Trees, and Sequences**, Cambridge (New York) 1997.
38. Harborne, J. B., C. A. Williams, *Anthocyanins and Flavonoids*, Nat Prod Rep 2001 Vol 18 pp. 310-333. http://www.rsc.org/ej/NP/1998/a815631y.pdf
39. Harborne, *Spectral Methods of Characterizing Anthocyanins*, Biochemical Journal, pp 22-28, 1958. http://www.biochemj.org/bj/070/0022/0700022.pdf
40. Harvey, P., M. Pagel, The Comparative Method in Evolutionary Biology, Oxford (New York) 1991.
41. Hasegawa, M., T. Yahara, Bimodal Distribution of Flowering Time in a Natural Hybrid Population of Daylily (Hemerocallis fulva) and nightly (Hemerocallis citrina), Journal of Plant Research, 2006 pp 63-68.
42. Hatzimanikatis, V., *Dynamical Analysis of Gene Networks Requires Both mRNA and Protein Expression Information*, Metabolic Engr, Vol 1, 1999, pp. 275-281.
43. Haywood, V. et al, *Plasmodesmata: Pathways for Protein and Ribonucleoprotein Signaling*, The Plant Cell, 2002 PP 303-325.
44. Hildebrand, F. B., **Numerical Analysis**, 2nd Edition, Dover (New York) 1987.
45. Holton, T., E. Cornish, *Genetics and Biochemistry of Anthocyanin Biosynthesis,* The Plant Cell, Vol 7, 1995, pp 1071-1083.
46. Innan, H., et al, *A Method for Estimating Nucleotide Diversity from AFLP Data*, Genetics, Vol 151 March 1999, pp. 1157-1164.
47. Jaakola, L. et al, *Expression of Genes Involved in Anthocyanin Biosynthesis*, Plant Physiology, Vol 130 Oct 2002, pp 729-739.
48. Jenkins, F. A., H. E. White, **Fundamentals of Optics**, McGraw Hill (New York) 1957.
49. Judd, D. B. et al, **Color** 2ND Edition, Wiley (New York) 1963.
50. Judd, W., et al, **Plant Systematics**, 3rd Ed, Sinauer (Sunderland, MA) 2008.
51. Judd, W., R. Olmstead, A Survey of Tricolpate (Eudicot) Phylogenetic Relationships, Journal of Botany Oct 2004, pp 1627-1644.
52. Kadar, S., et al, *Modeling of Transient Turing Type Patterns in the Closed Chlorine Dioxide-Iodine-Malonic Acid-Starch Reaction System*, J Phys Chem, Vol 99 1995 pp. 4054-4058.
53. Kang, S., M. Chung, Hemerocallis taeanensis (Liliaceae) a New Species from Korea, Systematic Botany 1997 Vol 22 pp 427-431.
54. Kang, S., M. Chung, Genetic Variation and Population Structure in Korean Endemic Species, IV Hemerocallis hakuunensis, Journal of Plant Research, Vol 110 1997 pp 209-217.
55. Kang, S., M. Chung, High Levels of Allozyme Variations Within Populations and Low Allozyme Divergence within and Among Species, American Journal of Botany, Vol 87 2000 pp 1634-1646.
56. Kohane, I., et al, **Microarrays for an Integrative Genomics**, MIT Press (Cambridge) 2003.
57. Koopman, W., et al, *AFLP Markers as a Tool to Reconstruct Complex Relationships: A Case Study in Rosa*, Journal of Botany, Vol 95 2008 pp 353-366.
58. Lee, D., **Nature's Palette**, University of Chicago Press (Chicago) 2007.
59. Lesk, A., **Bioinformatics**, Oxford (New York) 2002.
60. Levi, L., **Applied Optics**, Wiley (New York) 1968.
61. Mauseth, J. D., **Plant Anatomy**, Benjamin (Menlo Park, CA) 1988.
62. Mayo, O., **The Theory of Plant Breeding**, Oxford (New York) 1987.
63. Mayr, E., **Evolution and the Diversity of Life,** Harvard University Press (Cambridge) 1976.
64. Mayr, E., **Population, Species and Evolution**, Harvard University Press (Cambridge) 1970.
65. Mayr, E., **The Growth of Biological Thought**, Harvard University Press (Cambridge) 1982.
66. Mayr, E., **Toward a New Philosophy of Biology**, Harvard University Press (Cambridge) 1988.
67. McGarty, T. P., *Flower Color and Means to Determine Causal Anthocyanins And Their Concentrations*, MIT 2008, http://www.telmarcgardens.com/Documents%20Papers/Flower%20Color%20and%20Means%20to%20Determine%2002.pdf
68. McGarty, T. P., *On the Structure of Random Fields Generated by a Multiple Scatter Medium*, PhD Thesis, MIT 1971. http://mit.edu/mcgarty/www/MIT/Paper%20Hypertext/1971%20PhD%20MIT.pdf

69. McGarty, T., *Gene Expression in Plants*: Use of System Identification for Control of Color, MIT, 2007. http://mit.edu/mcgarty/www/MIT/Paper%20Hypertext/2007%20Gene%20Expression%20IEEE%2007%2002.pdf .
70. McGarty, T., **Stochastic Systems and State Estimation**, Wiley (New York) 1974.
71. McMurry, J., Begley, T., The **Organic Chemistry of Biological Pathways**, Roberts & Company Publishers, 2005.
72. Milgrom, L. R**., The Colours of Life**, Oxford (New York) 1997.
73. Mohr, H., P. Schopfer, **Plant Physiology**, Springer (New York) 1995.
74. Mol, J, et al, *How Genes Paint Flowers and Seeds, Trends in Plant Science*, Vol 3 June 1998, pp 212-217.
75. Mol, J., et al, *Novel Colored Plants*, Current Opinion in Biotechnology, Vol 10, 1999, pp 198-201.
76. Mueller, U., L. Wolfenbarger, AFLP Genotyping and Fingerprinting, Trends Ecol. Evol. **14: 1999, pp** 389–394. .
77. Munson, R., **Hemerocalis, The Daylily**, Timber Press (Portland, OR) 1989.
78. Murray, J., **Mathematical Biology**, Springer (New York) 1989.
79. Murrell, J., *Understanding Rate of Chemical Reactions*, University of Sussex.
80. Naik, P. S., et al, *Genetic manipulation of carotenoid pathway in higher plants*, Current Science, Vol 85, No 10, Nov 2003, pp 1423-1430.
81. Nei, M., S.. Kumar, **Molecular Evolution and Phylogenetics**, Oxford (New York) 2000.
82. Nicklas, K. J., *The Bio Logic and Machinery of Plant Morphogenesis*, Am Jour Bot 90(4) 2003 pp. 515-525.
83. Niklas, K. J., **The Evolutionary Biology of Plants**, Chicago (Chicago) 1997.
84. Noguchi, J., H. De-yuan, *Multiple origins of the Japanese nocturnal Hemerocalis citrina*, Int Jrl Plant Science, 2004, Vol 16, pp. 219-230.
85. Norton, J. 1972. *Hemerocallis* Journal 26 (3) in Bisset, K. 1976. *Spectrophotometry, Chromatography and Genetics of Hemerocallis Pigments*. Dissertation, Florida State Univ.
86. Norton, J., *Some Basic Hemerocallis Genetics*, American Hemerocallis Society, 1982.
87. Oparka, K. J., A. G. Roberts. Plasmodesmata*, A Not So Open and Shut Case*, Plant Phys, Jan 2001, Vol 125 pp. 123-126.
88. Percus, J., **Mathematics of Genome Analysis**, Cambridge (New York) 2002.
89. Perkins, T., et al, *Inferring Models of Gene Expression Dynamics*, Journal of Theoretical Biology, Vol 230, 2004, pp. 289-299.
90. Petit, T. *The Patterned Daylily*, The Daylily Journal, Vol 62 No 2 2007 pp. 125-141.
91. Petit, T L, J P Peat, **The New Encyclopedia of Daylilies**, Timber (Portland) 2008.
92. Planet, P. et al, Systematic Analysis of DNA Microarray Data: Ordering and Interpreting Patterns of Gene Expression, Cold Spring Harbor, Genome Research, 2001 pp 1149-1155.
93. Rossi, B., **Optics**, Addison Wesley (Reading, MA) 1957.
94. Saitou, N., M. Nei, *The Neighbor Joining Method: A New Method for Reconstructing Phylogenetic Trees,* Molecular Biological Evolution, Vol 4 1987 pp 406-425.
95. Sansone, G et al, **Orthogonal Functions**, Academic Press (Mew York) 1958.
96. Sattath, S., A. Tversky, Additive Similarity Trees, Psychometrica, Vol 42, 1977 pp. 319-345.
97. Schnell. S, T. Turner, *Reaction Kinetics in Intracellular Environments with Macromolecular Crowding*, Biophys and Molec Bio vol 85 2004 pp. 235-260.
98. Sears, F. W., **Optics**, Addison Wesley (Reading, MA) 1949.
99. Sokal, R., P. Sneath, **Principle of Numerical Taxonomy**, Freeman (San Francisco) 1963.
100. Soltis, D. et al, *Evolution of Genome Size in Angiosperms*, Journal of Botany, Nov 2003, pp 1596-1603
101. Somasundaram, S., M. Kalaiselvam, Molecular Tools for Assessing Genetic Diversity, UN University Course, http://ocw.unu.edu/international-network-on-water-environment-and-health/unu-inweh-course-1-mangroves/Molecular_Tools__for_Assessing_Genetic_Diversity.pdf .
102. Stace, C., **Plant Taxonomy and Biosystematics,** Arnold (London) 1989.
103. Stout, A.B., **Daylilies**, Saga Press (Millwood, NY) 1986.
104. Stout, A., The Inflorescence in Hemerocallis, Bulletin of the Torrey Club, Vol 68 1941 pp 305-316.

105. Stout, A., C. Chandler, Pollen Tube Behavior in Hemerocallis with Special Reference to Incompatibilities, of the Torrey Club, Vol 60 1933 pp 397417.
106. Strickberger, M., **Genetics**, 2nd Ed McMillan (New York) 1976.
107. Strong, J., **Concepts of Classical Optics**, Freeman (San Francisco) 1958.
108. Studier, J., K. Kappler, *A Note on the Neighbor Joining Algorithm*, Molecular Biological Evolution, Vol 5 1988 pp 729-731.
109. Stuessy, T., **Case Studies in Plant Taxonomy**, Columbia University Press (New York) 1994.
110. Szallasi, Z. **System Modeling in Cellular Biology: From Concepts to Nuts and Bolts**. MIT Press (Cambridge) 2006.
111. Taiz, L., E. Zeiger, **Plant Physiology**, Benjamin Cummings (Redwood City, CA) 1991.
112. Taubes, C. H., Modeling Differential Equations in Biology, Cambridge (New York) 2001.
113. Tetter, A., et al, Primer on Medical Genomics Part III, Mayo Clinic Proc, Vol 77 2002 pp 927-940.
114. Tinoco, I. et al, **Physical Chemistry**, Prentice Hall (Englewood Cliffs, NJ) 1995.
115. Tobias, A., *Directed Evolution of Biosynthetic Pathways to Carotenoids with Unnatural Carbon Bonds*, PhD Thesis, Cal Tech, 2006.
116. Tomkins, J. R., *DNA Fingerprinting in Daylilies*, Parts I and II, Daylily Journal, Vol 56 No 2 and 3 2001, pp. 195-200 and pp. 343-347.
117. Tomkins, J. R., *How much DNA is in a Daylily*, Daylily Journal, Vol 58 No 2 2003, pp. 205-209.
118. Tomkins, J., et al, *Evaluation of genetic variation in the daylily (Hemerocalis) using AFLP markers*, Theor Appl Genet Vol 102, 2001, pp. 489-496.
119. Turing, A., *The Chemical Basis of Morphogenesis*, Phil Trans Royal Soc London B337 pp 37-72, 19459.
120. Van Trees, H. L., **Detection, Estimation and Modulation Theory**, Wiley (New York) 1968.
121. Vohradsky, J., *Neural Network Model of Gene Expression*, FASEB Journal, Vol 15, March 2001, pp. 846-854.
122. Vos, P., et al, *AFLP: A New Technique for DNA Fingerprinting, Nuclear Acids Research*, Vol 23 1995 pp 4407-4414.
123. Wade, L. G., **Organic Chemistry**, Prentice Hall (Saddle River, NJ) 2003.
124. Watson, J. et al, **Recombinant DNA**, 3rd Ed, Freeman (New York) 2007.
125. Watson, J., et al, **Molecular Biology of the Gene**, Benjamin Cummings (San Francisco) 2004.
126. Weir, B., **Genetic Data Analysis**, Sinauer (Sunderland, MA) 1990.
127. Wen, X., et al, *Large Scale Temporal Gene Expression Mapping of Central Nervous System Development*, Proc Nat Acad Sci Neurology 1998 Vol 95 pp 334-339.
128. Winkel-Shirley, B., *Flavonoid Biosynthesis*, Plant Physiology, Vol 126 June 2001 pp 485-493. http://www.plantphysiol.org/cgi/reprint/126/2/485
129. Yasumoto, A., T., Yahara, Reproductive Isolation on Interspecific Backcross of F1 Pollen to Parental Species Hemerocallis fulva and Hemerocallis citrina, Journal of Plant Research, Vol 2008, pp .
130. Yu, O. et al, *Flavonoid Compounds in Flowers: Genetics and Biochemistry*, General Science Books, http://www.danforthcenter.org/yu/pdf/e-flower-2006.pdf
131. Zambryski, P., C*ell to Cell Transport of Proteins and Fluorescent Tracers via Plasmodesmata during Plant Development*, Jour Cell Bio Vol 161 No 2 Jan 2004, pp. 165-168.