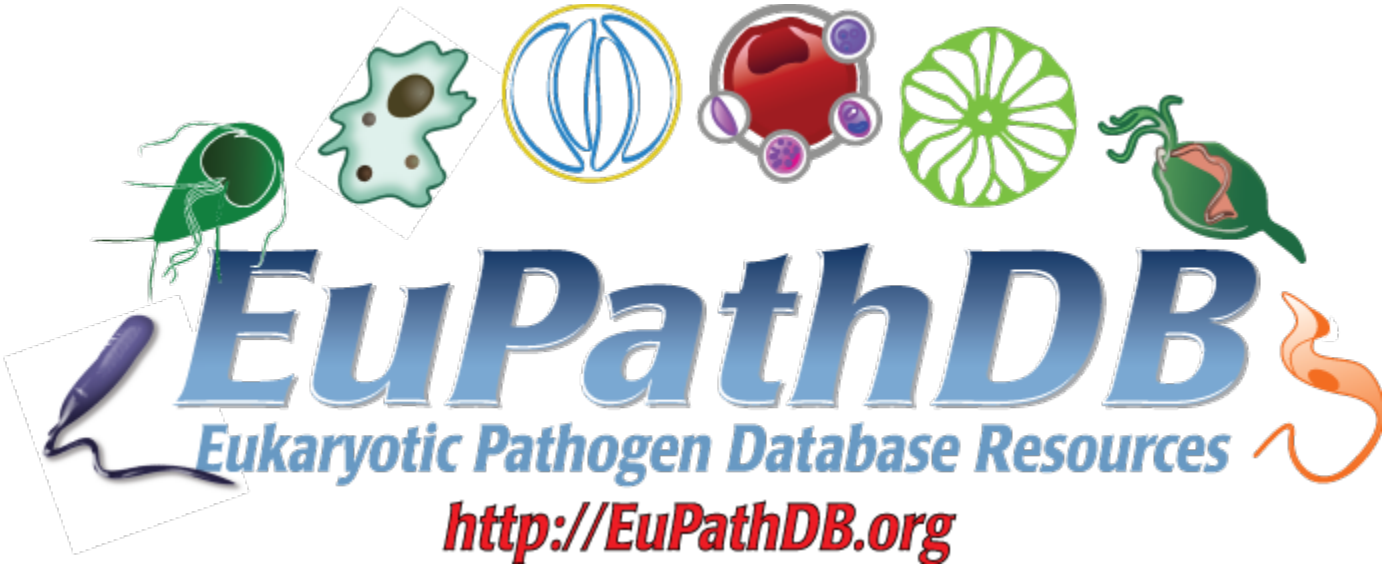
Welcome!

EuPathDB Workshop 2013



Instructors & Friendly Faces

- Betsy Wenthe

- Brian Brunk

- Susanne Warrenfeltz

- Cristina Aurrecoechea

- Jessie Kissinger

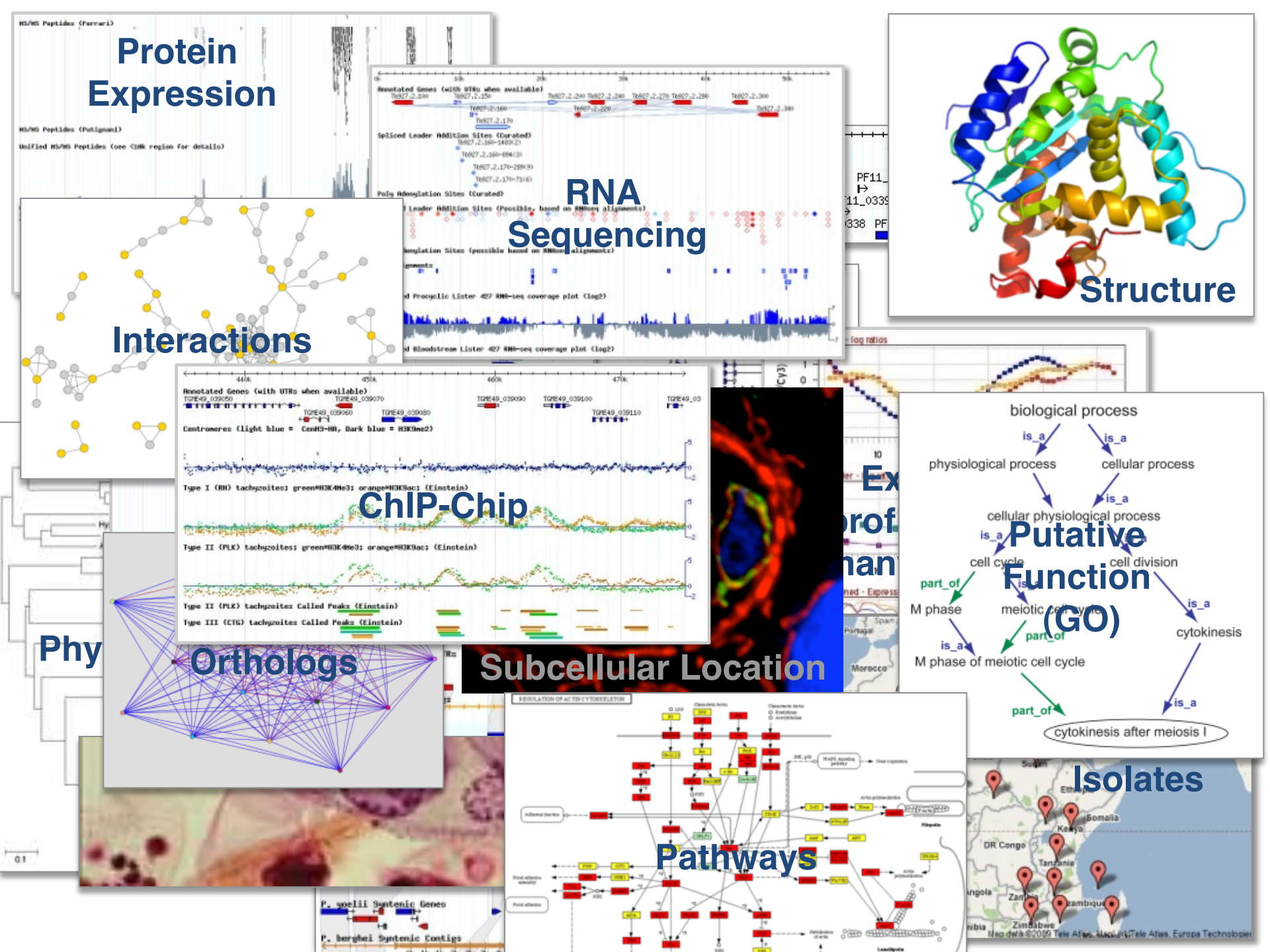
- Omar Harb

- Eileen Kraemer

Crash Course in Omics Terminology, Concepts & Data Types

Jessie Kissinger

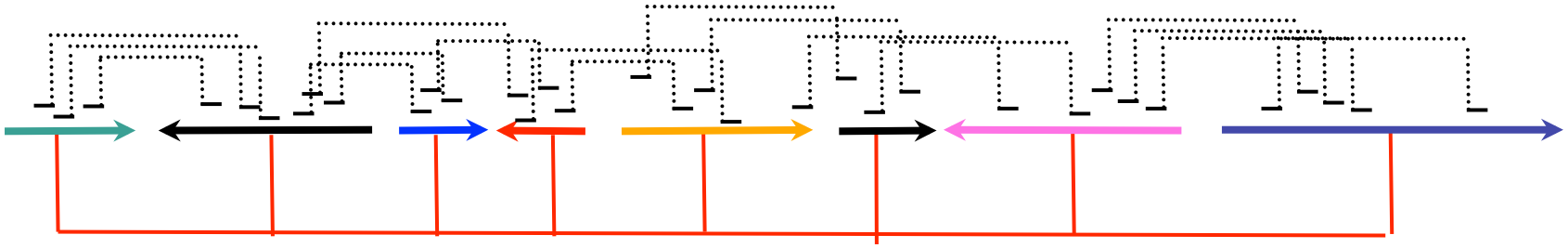
June 2, 2013


Protein Expression
RNA Sequencing
Structure
Interactions
ChIP-Chip
Putative Function (GO)
biological process
is_a
physiological process
cellular process
cellular physiological process
cell cycle
cell division
part_of
M phase
meiotic cell cycle
cytokinesis
M phase of meiotic cell cycle
cytokinesis after meiosis I
Orthologs
Subcellular Location
Pathways
Isolates
Somalia
Kenya
DR Congo
Tanzania
Zambia
Zimbabwe
P. woelli Syntenic Genes
P. berghei Syntenic Contigs

Genome assembly

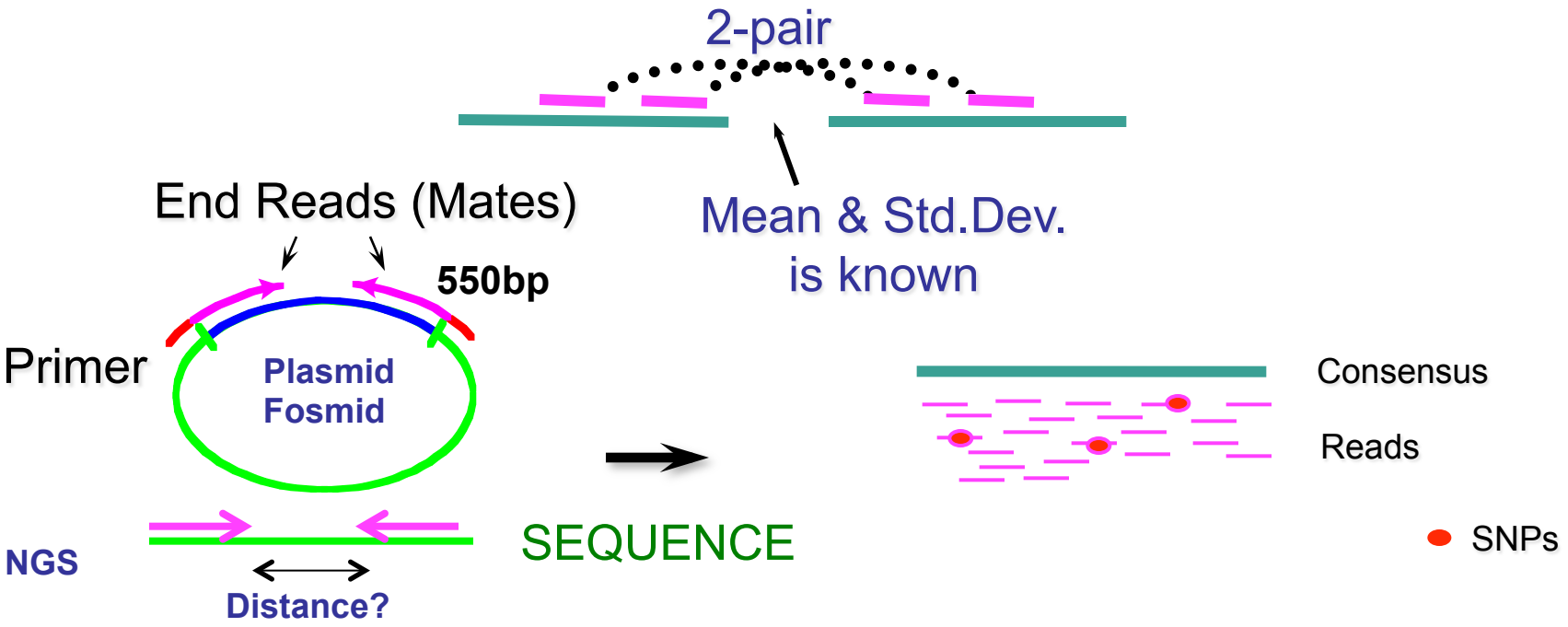
- 5X random genome shotgun
- Library insert size
- Paired-end? (Mated end pairs)
- Contigs
- Scaffolds

Pairs Give Order & Orientation

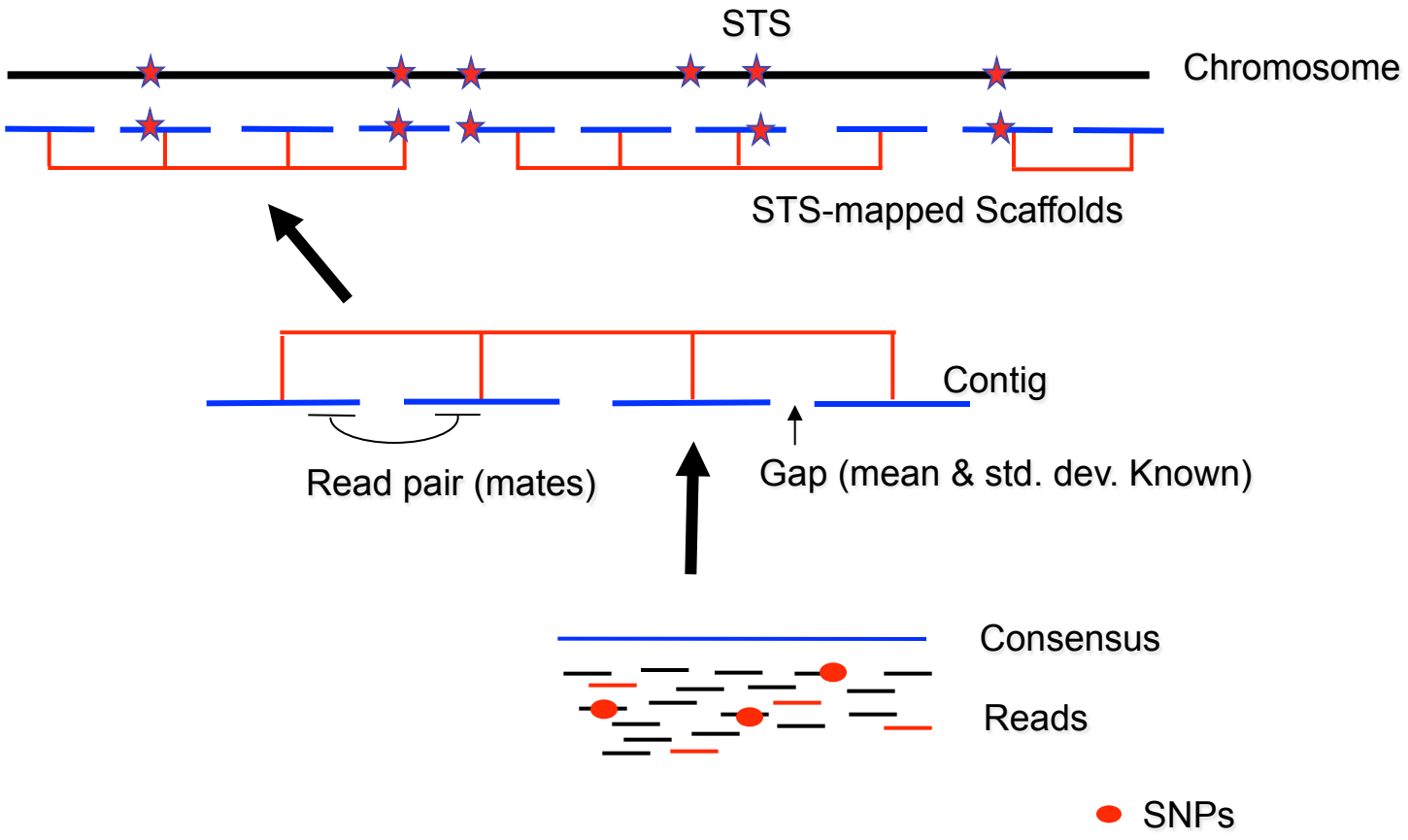


Scaffold

Gaps in scaffolds are traditionally indicated by 100 "N's"

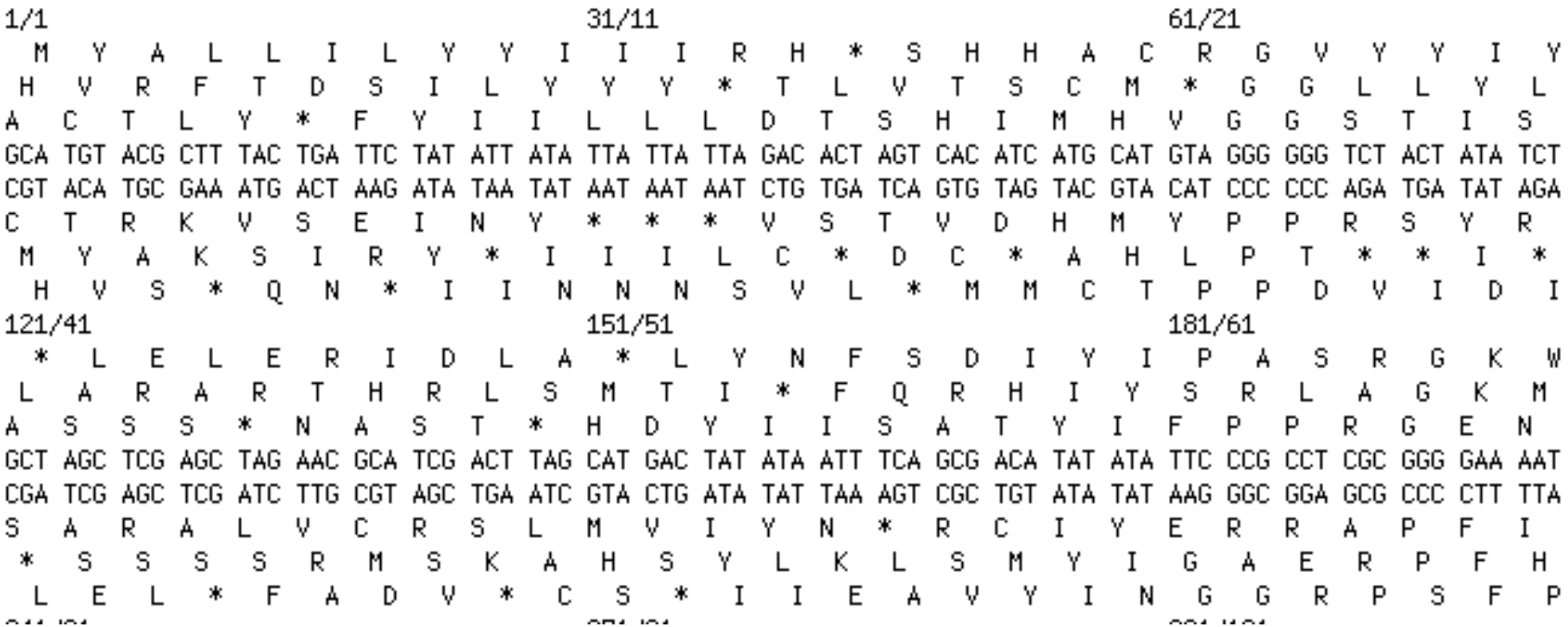


Anatomy of a WGS Assembly



AAGCTTCGCCAGGCTGTAAATCCCGTGAGTCGTCCTCACAAATCATCAAGCAGGTGTCCTCAGGGAGACTGCCTGACTGAGTTATGCTAATTCCTTTCTACTTTGGCGTGGTCACGTGTA
ACCATATCCGAATCATTTCTCTAGCCCTACGAACAGGTAAGAGCGCTAGGGATGTCCGTGGAGTAGTGTGCTTACTCGATAATATTCAGTTGGGACTACCAGCGAGGCGCTCGCTTTGCT
CACGCAATGCCTGAGACAGTTGCAGAATGAATGGTAACCGACAAACGCGTTCATATGCGTTTTCAAACTTAGTAGACGCGTACTGTCTGAAACTGGCGGTCACAGGCACCAGATAACGCC
CTTGGCATCGGCATGTCTCGTACAGAGGTCCGTATGTAGTGCCACGACTTCTAAATCCGGCGACAGGCTGGTCTTTTGTCTTACCACGTATTAGCCCGCGTGCGATTTCTCGGAGCGCAC
CTGTTCAACACTAGAAAACGGAGTTTCCTGATCGAGAAGCCACCACCTTTCCAGAAGTTGAACGCTAGCATGTCATTCGATTTTCACCCCCCGCGTAGTTCCTGTGTGTCATTCGTTGTC
GAGACAACTCTGTCCCGCCCCGGTGCTGTTCCATATGCGTGACTTTCCCGCAATTTTTTCAGACTTTCAGGAAAGACAGGCTCCGGAACGATCTCGTCCATGACTGGTAAATCCACGACA
CCGCAATGGCCCCCAGCACCTCTATCTCTCGTGCCAGGGGACTAACGTTGTATGCGTCTGCGTCTTGTCTTTTTGCATTCGCTTTCCAAAAAAGAGAGCCATCCGTTCCCCCGCACATTC
AACGCCGCGAGTGCGGTTTTTGTCTTTTTTGAGTGGTAGGACGCTTTTCATGCGCGAACTACGTGGACATTAAGTTCCATTCTCTTTTTCGACAGCACGAAACCTTGCATTCAAACCCGC
CCGCGGAAGATCCGATCTTGCTGCTGTTCGCAGTCCCAGTAGCGTCCTGTCGGCCGCGCCGTCTCTGTTGGTGGGCAGCCGCTACACCTGTTATCTGACTGCCGTGCGCGAAAATGACGC
CATTTTTGGGAAAATCGGGGAACTTCATTCTTTAAAAGTATGCGGAGGTTTCCTTTTTCTTCTGTTCGTTTCTTTTTCTCGGGTTTGATAACCGTGTTCGATGTAAGCACTTTCCGTCTC
TCCTCCGTGCTTTGTTCGACATCGAGACCAGGTGTGCAGATCCTTCGCTTGTCGATCCGGAGACGCGTGTCTCGTAGAACCTTTTCATTTTACCACACGGCAGTGCGGAGCACTGCTCTG
AGTGCAGCAGGGACGGGTGAAGTTTCGCTTTAGTAGTGCGTTTCTGCTCTACGGGGCGTTGTCGTGTCTGGGAAGATGCAGAAACCGGTGTGTCTGGTCGTCGCGATGACCCCCAAGAGG
GGCATCGGCATCAACAACGGCCTCCCGTGGCCCCACTTGACCACAGATTTCAAACACTTTTCTCGTGTGACAAAAACGACGCCCGAAGAAGCCAGTCGCCTGAACGGGTGGCTTCCCAGG
AAATTTGCAAAGACGGGCGACTCTGGACTTCCCTCTCCATCAGTCGGCAAGAGATTCAACGCCGTTGTCATGGGACGGAAAACCTGGGAAAGCATGCCTCGAAAGTTTAGACCCCTCGTG
GACAGATTGAACATCGTCGTTTCCTCTTCCCTGTGAGCACACACAGTAGTCGCCACACGCTGTTTGAGACGTGTCAATCTCCAAGAGTGTGGACGCTGTTCCACGTCTTCAAATGTTTCC
CAACATCCGTCGTCTAGTAGACACACCAACAAAAAGCACACGGCGAATCTGCTCATCGGAGGGAGGAGCCGGGGGGCACACAACTATCCTCAACTCTCGAACGAACATATCCGGGGCCGC
GAAGACGTCCAGTCTCTCAAATCCAACCCGGAACGCAAACATTTCTGCATCAAGTCACGATTGCGCCGGTACCTCCATGTGTAAGCAGTTCCATGAAACCTCCGATATTACACACGACTG
TGGATATGAATTATATGCAGATGCATATATACTGAGACGCCGATGCAACTATAGGTTTCCTGGCCCTCCATGGATATTTCAGACCTTCCTCTCACATTTGGTTTGCCCGTACACCTCCGT
TACGCTTTTTTTCTGGCTTTCTTCTTCGTCTCTGTTTATCAGCAAAGAAGAAGACATTGCGGCGGAGAAGCCTCAAGCTGAAGGCCAGCAGCGCGTCCGAGTCTGTGCTTCACTCCCAGC
AGCTCTCAGCCTTCTGGAGGAAGAGTACAAGGATTCTGTCGACCAGATTTTTGTCGTGGGTATGTTGTCCTAAACTCCTTGGAACTCCATTCTTGGTCAGAAACGTACTGAAACTGTATA
CATGTATATACAGATGTATGGATAATATCTAGAGAAGATACAGGGAAGACTGGCAAGGATGAAAAGACATGCAGCTTTAACGAAGCAGAGGGCATTGGCGAGAGGGACGCCCGTTATGCT
GTGTGATGTGGCTGTGAATCTTACCTCGCCGTTTGACTTGCTGCAGCGCTTTGTCCACTTGAACGTGACTTCTTGTTTCTACCTTCCCCAACGCCTTCTATTCCCTTCACTGCGAAAGCG
CGCTCAGTGGGCCGTCACCGAACACCCTTGGTTCTTTCGTTCAGCTGTTGTCCTCTTTCTCGCGTTGCTTCCTGTGGCGTCGTGGCTCGGCTTCTCTCTCTTTCCTGTTGGTGCGTCCAG
ACTATGTCGCCTGTTTCCCCACCCTTCTCGGCTTGTGCTTTCAGGAGGAGCGGGACTGTACGAGGCAGCGCTGTCTCTGGGCGTTGCCTCTCACCTGTACATCACGCGTGTAGCCCGCGA
GTTTCCGTGCGACGTTTTCTTCCCTGCGTTCCCCGGAGATGACATTCTTTCAAACAAATCAACTGCTGCGCAGGCTGCAGCTCCTGCCGAGTCTGTGTTCGTTCCCTTTTGTCCGGAGCT
CGGAAGAGAGAAGGACAATGAAGCGACGTATCGACCCATCTTCATTTCCAAGACCTTCTCAGACAACGGGGTACCCTACGACTTTGTGGTTCTCGAGAAGAGAAGGAAGACTGACGACGC
AGCCACTGCGGAACCGGTAAGAGGCAACCGAAGCGCGTAGATAAGAAAAACAACAAAGAGAAGGTGAAACACGAAGAGAAGGGAAAATGCGGAGAAACCGTGGATTTACAAAGATATCAA
GAGCAATGCTTTGTGGAGATTTTTTTTAATTCAGTAGAGACACCCGCCGTGCGAGGTGTGTAGAAATAACTGCGACCCTGGAGACAGAGATGCCGCGAGTACACCACTTGTCGTTTTTCC
TCCTATGTTCATGACGGGTGCTGAACGTCTATCGTACTTAATTGGAGGAGTCGTCTCCGAAGCAGCTTTGGCTGGCCATCCGTGTGTTTGCCTTGTTCCTGAAAAGCCAGAAGGCGCTCC
ACAGTGAGGCGATATACAGGGACGCCTACCGGAGCCCCGTTTTCTGCCTTTGTCGACTCTTGCAGAGCAACGCAATGAGCTCCTTGACGTCCACGAGGGAGACAACTCCCGTGCACGGGT
TGCAGGCTCCTTCTTCGGCCGCAGCCATTGCCCCGGTGTTGGCGTGGATGGACGAAGAAGACCGGAAAAAACGCGAGCAAAAGGAACTGATTCGGGCCGTTCCGCATGTTCACTTTAGAG
GCCATGAAGAATTCCAGTACCTTGATCTCATTGCCGACATTATTAACAATGGAAGGACAATGGATGACCGAACGGGTAACGGCGACTGCGAGAAAAAGCCACACCGTTTTCTCCTGTGAT
TCTGTCCGCAAGCCCTCTTTTGCTTCATCCACCCTTTGCTATTCTCCGCCGCCTTCCTTTTCTGCTCCATGTTCAATTCGTTCGCTTCTTCAGTCTTTCCATCTTCCCCTGTTACCTCTG
TCATTCGTTTTCTTGCCTCTATTTAACTGTGTTCTACTCACAGTCTGCATTCCGCGATAGACGAGCTTCCACGTCTTGCGTCTCGACAAGCAACTGTCATTTGTACGCGCCTCCCTCCAC
CGTGAATCGGATTGTCGGTTCGCCGGTTCCTGGGTCAGAAAAGGCCTGCGCCAGTATTCTGAATAATACCCTTCGCCATTGTAAAGAGGCGAAGGAACAAAGAGATATTTCGGCGCATCT
TTTGTGCGGCGCGTTTCCTCGTGCTTCACACCGATGCCCTTCTGTGCATGTCTTCTGCTCCTCGTCCTTCTCTCTTTTTCCCTGTTTAGGCGTTGGTGTCATCTCCAAATTCGGCTGCAC
TATGCGCTACTCGCTGGATCAGGCCTTTCCACTTCTCACCACAAAGCGTGTGTTCTGGAAAGGGTAAGGGCGTCTTCAGTGAATGCATATATTTGACTTCAGACATTCTTAACTGTTTGA
CAACCAACGTACAAATTTGTTTGTCCGTGTGCGTGTTCGACATGTCAAGTATGTGAAGAGTCGCTACTGTAGACTAACGCACGAACCAGATTTGTTTATCTGCATGCGCTGTGCACCCGT
TTCTGAGTGTCTGGAGTTTCCGCAACCTTCCTTTGAATTTCTGGGTTCGTTTTTTTATGCGCGCACTGGTTTGCATGTGGCCTGAGAGAGCACAGATCGAAGGTGGGGTGATGTGGCGTC
GCTGCAGAGAAACTCCGGCGAAGGCGACAGATAAAGGAGAGTGGAAATCATTGAACAGTGTCGGTCGTCTGTTGTTTCGCAGGGTCCTCGAAGAGTTGCTGTGGTTCATTCGCGGCGACA
CGAACGCAAACCATCTTTCTGAGAAGGGCGTGAAGGCAAGTCTACGTTGTACCTCTTGTCTCTGCCGAAGCTCAGATGTCTCCACGGCGTTGGTTTCTTTTCGTTTTTGCTTTCGTGGCA
TTACCATCGAGTCACCACTCATAGTTGCGTGTGTCTACATGTTTTCTAGAACGTCCGTTGTGTTGCCTCGTGGCGACCGGCGCGAGTGTATGTACCCTGCGCTGTCAGAAGTTGATCCTT

Six Frame Translation ORF-finding



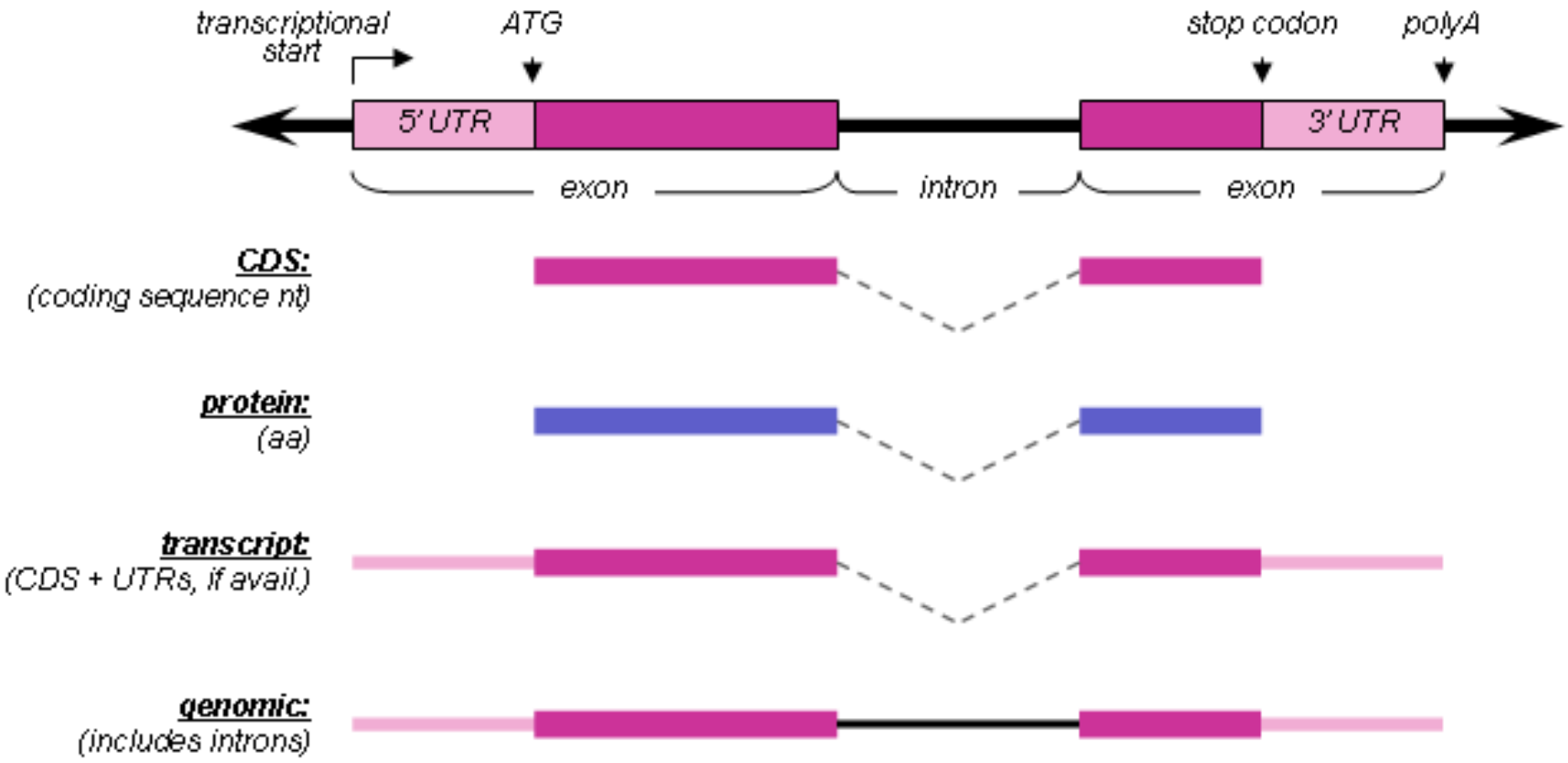
ORFs ≠ Genes

ATGCAGAAACCGGTGTGTCTGGTCGTCGCGATGACCCCCAAGAGGGGCATCGGCATCAACAACGGCCTCCCGTGGCCCC
ACTTGACCACAGATTTCAAACACTTTTCTCGTGTGACAAAAACGACGCCCGAAGAAGCCAGTCGCCTGAACGGGTGGCT
TCCCAGGAAATTTGCAAAGACGGGCGACTCTGGACTTCCCTCTCCATCAGTCGGCAAGAGATTCAACGCCGTTGTCATG
GGACGGAAAACCTGGGAAAGCATGCCTCGAAAGTTTAGACCCCTCGTGGACAGATTGAACATCGTCGTTTCCTCTTCCC
TCAAAGAAGAAGACATTGCGGCGGAGAAGCCTCAAGCTGAAGGCCAGCAGCGCGTCCGAGTCTGTGCTTCACTCCCAGC
AGCTCTCAGCCTTCTGGAGGAAGAGTACAAGGATTCTGTCGACCAGATTTTTGTCGTGGGAGGAGCGGGACTGTACGAG
GCAGCGCTGTCTCTGGGCGTTGCCTCTCACCTGTACATCACGCGTGTAGCCCGCGAGTTTCCGTGCGACGTTTTCTTCC
CTGCGTTCCCCGGAGATGACATTCTTTCAAACAAATCAACTGCTGCGCAGGCTGCAGCTCCTGCCGAGTCTGTGTTCGT
TCCCTTTTGTCCGGAGCTCGGAAGAGAGAAGGACAATGAAGCGACGTATCGACCCATCTTCATTTCCAAGACCTTCTCA
GACAACGGGGTACCCTACGACTTTGTGGTTCTCGAGAAGAGAAGGAAGACTGACGACGCAGCCACTGCGGAACCGAGCA
ACGCAATGAGCTCCTTGACGTCCACGAGGGAGACAACTCCCGTGCACGGGTTGCAGGCTCCTTCTTCGGCCGCAGCCAT
TGCCCCGGTGTTGGCGTGGATGGACGAAGAAGACCGGAAAAAACGCGAGCAAAAGGAACTGATTCGGGCCGTTCCGCAT
GTTCACTTTAGAGGCCATGAAGAATTCCAGTACCTTGATCTCATTGCCGACATTATTAACAATGGAAGGACAATGGATG
ACCGAACGG

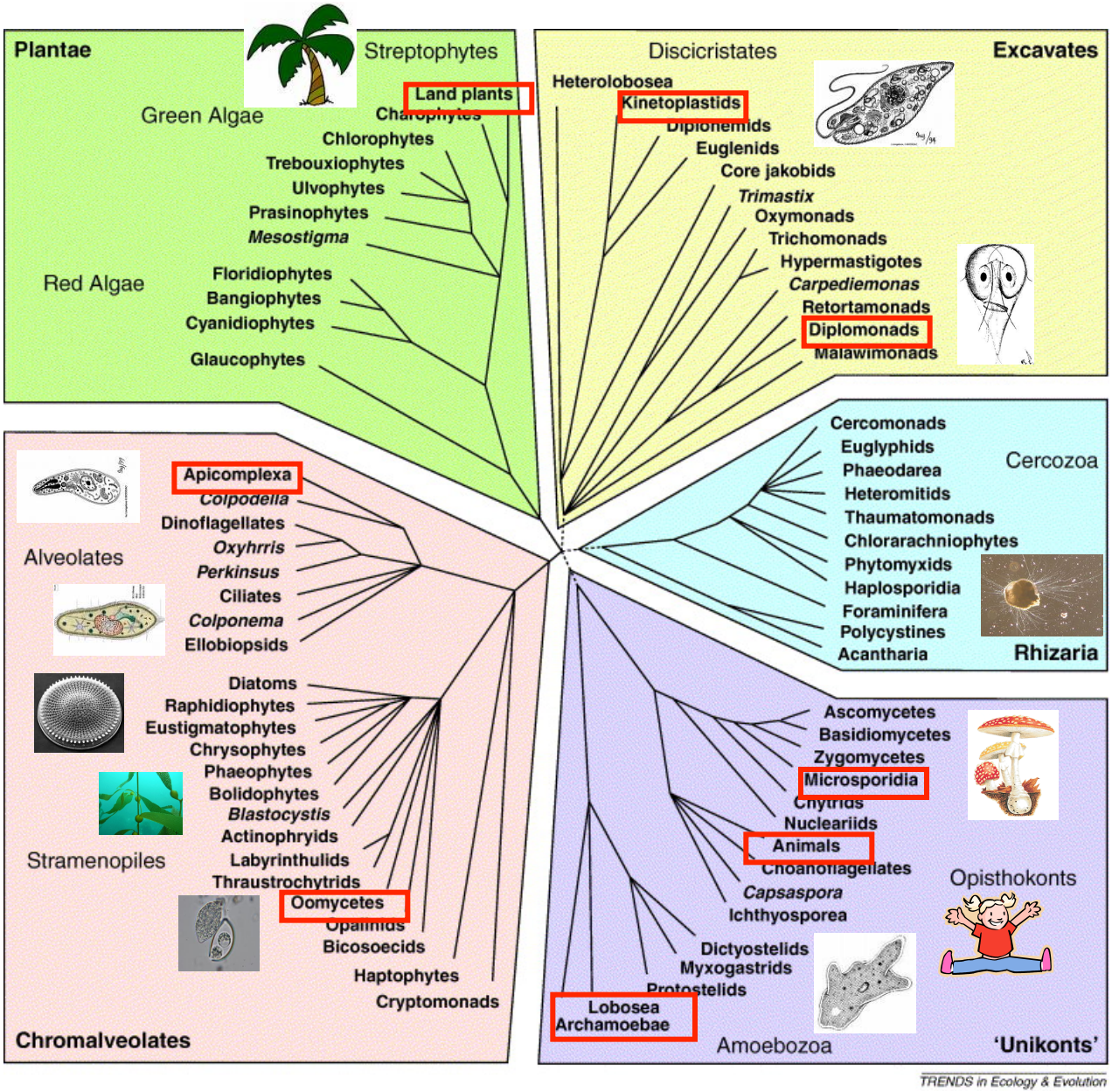
>Translation Frame 1

MQKPVCLVVAMTPKRGIGINNGLPWPHLTTDFKHFSRVTKTTPEEASRLN
GWLPRKFAKTGDSGLPSPSVGKRFNAVVMGRKTWESMPRKFRPLVDRLNI
VVSSSLKEEDIAAEKPQAEGQQRVRVCASLPAALSLLEEEYKDSVDQIFV
VGGAGLYEAALSLGVASHLYITRVAREFPCDVFFPAFPGDDILSNKSTAA
QAAAPAESVFVPFCPELGREKDNEATYRPIFISKTFSDNGVPYDFVVLEK
RRKTDDAATAEPSNAMSSLTSTRETTPVHGLQAPSSAAAIAPVLAWMDEE
DRKKREQKELIRAVPHVHFRGHEEFQYLDLIADIINNGRTMDDRT

Terminology

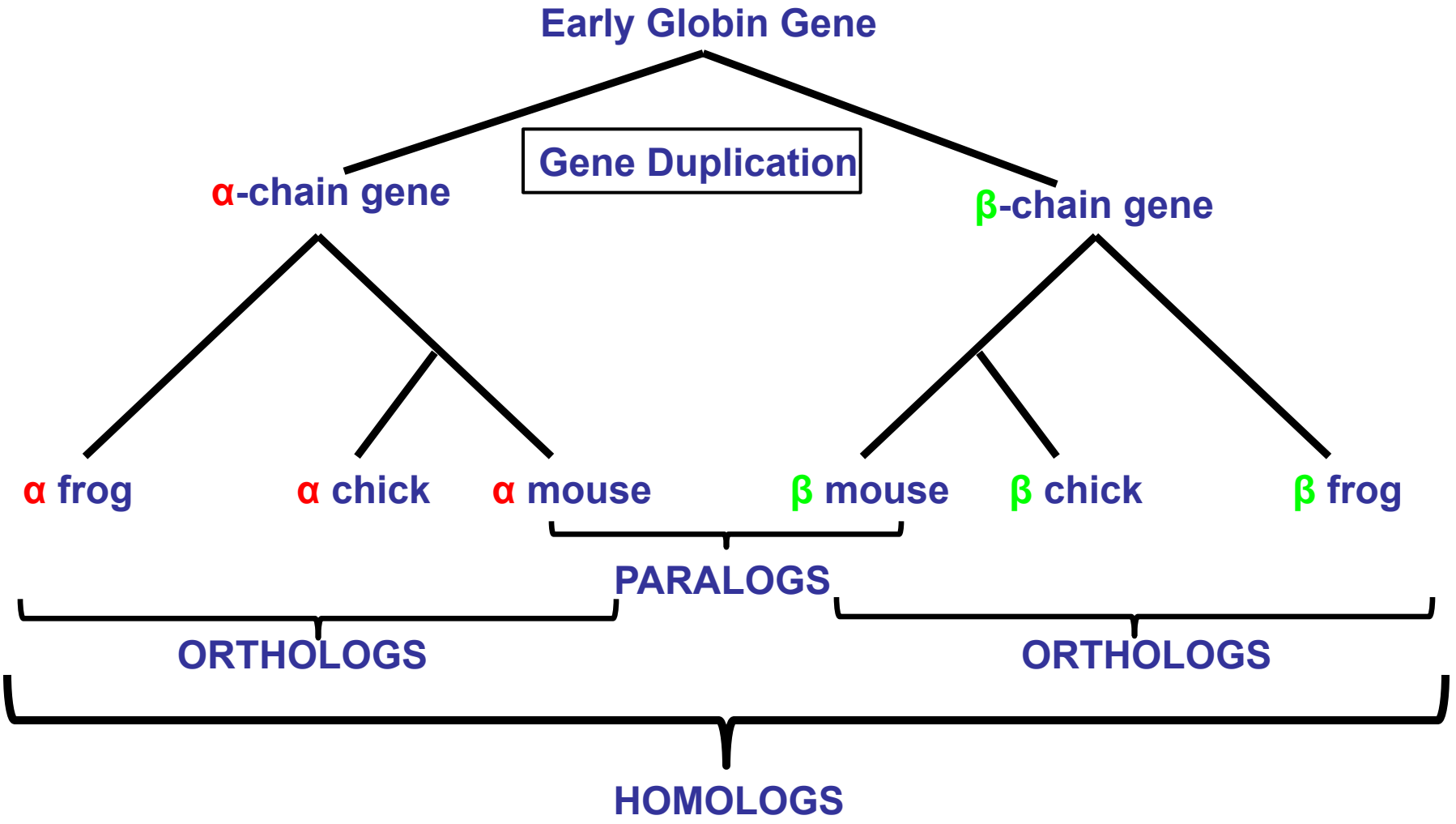


A Tree of Eukaryotes (Keeling, 2005)



TRENDS in Ecology & Evolution

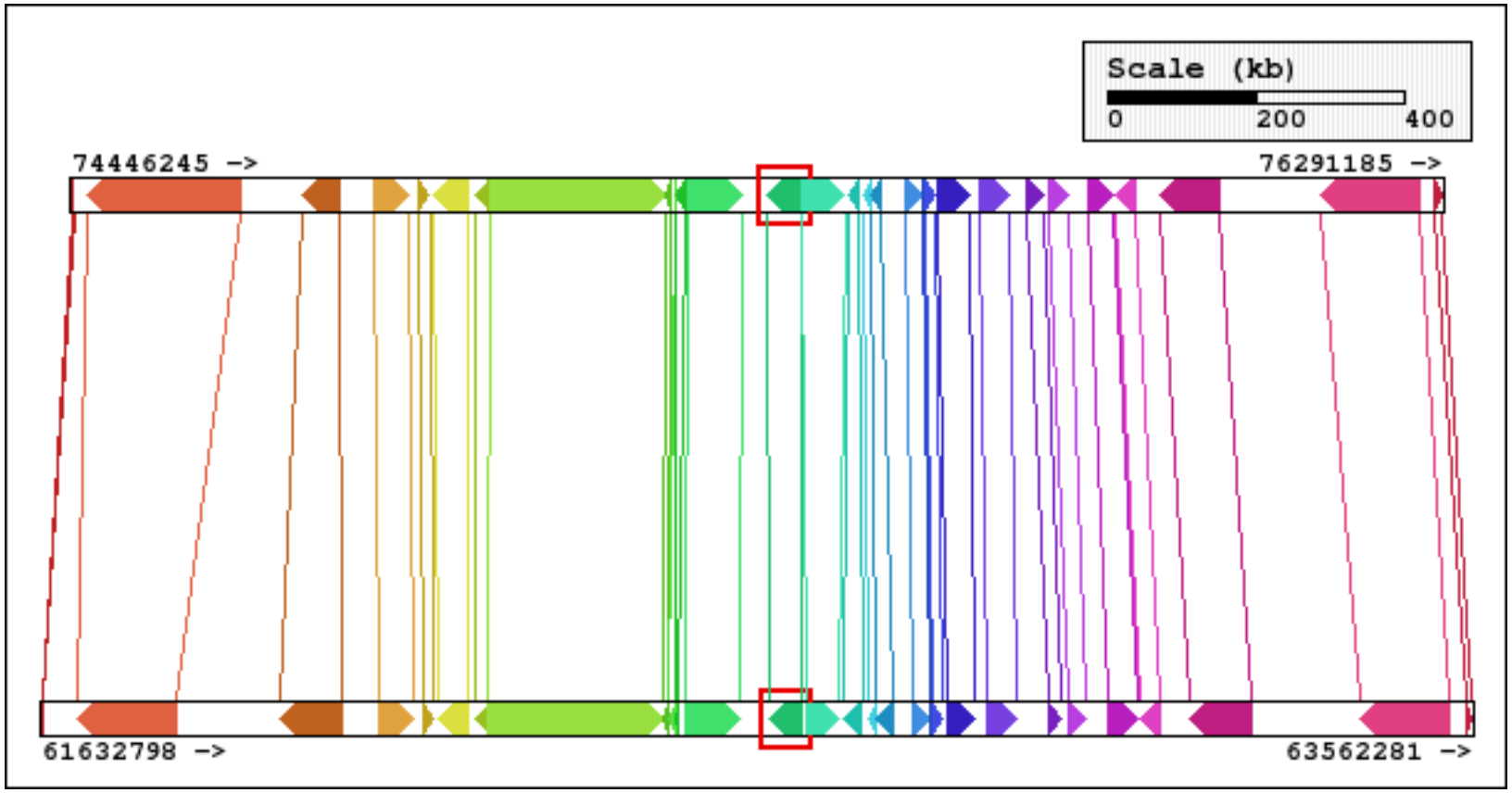
Homology


Early Globin Gene
Gene Duplication
α-chain gene
β-chain gene
α frog
α chick
α mouse
β mouse
β chick
β frog
PARALOGS
ORTHOLOGS
ORTHOLOGS
HOMOLOGS

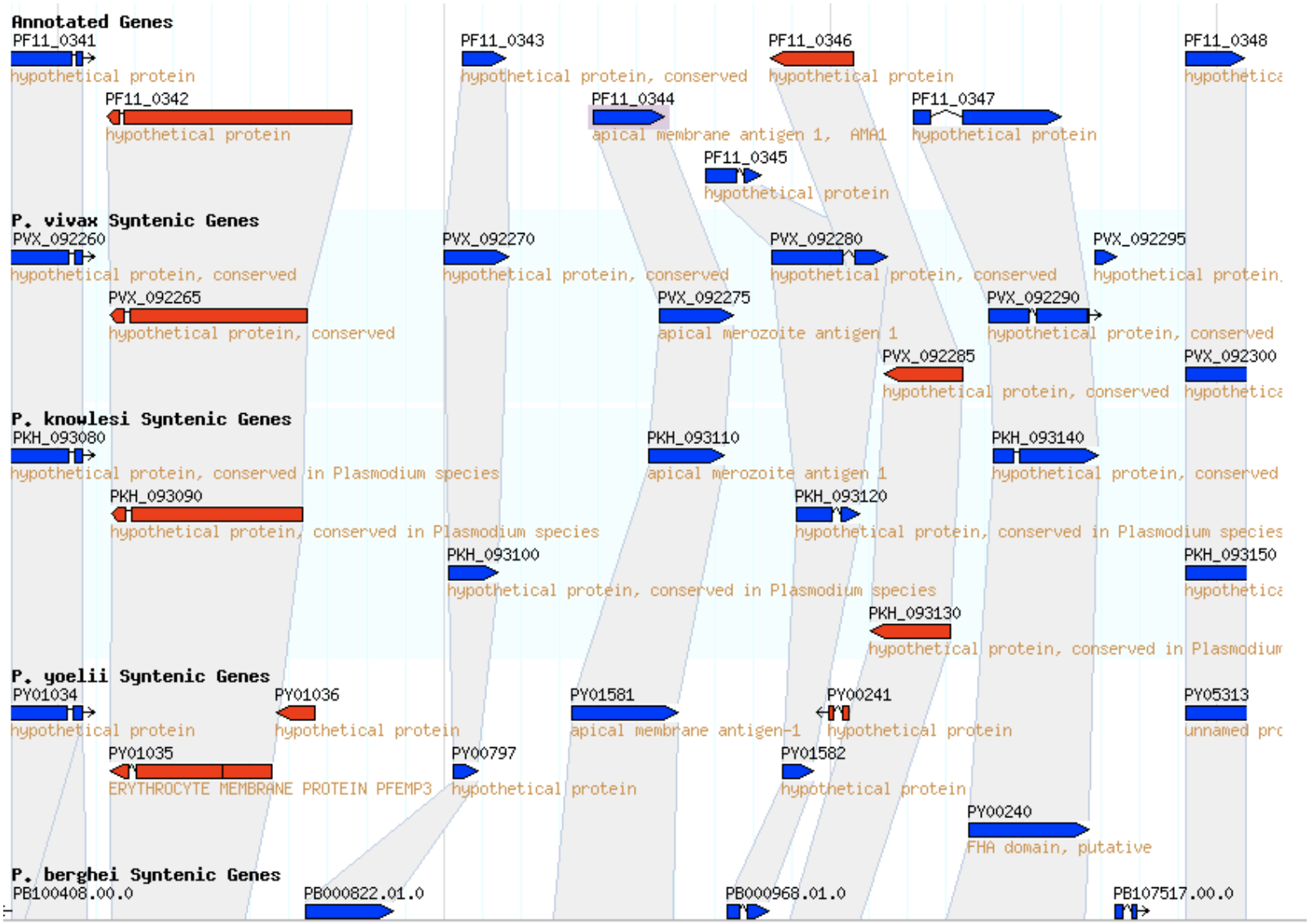
Evolutionary relationships

- Homology - related by evolutionary descent not equivalent to similarity
- Orthology - same gene in different organisms, e.g. alpha hemoglobin in humans and chimps
- Paralogy - genes within an organism related by gene duplication, e.g. alpha and beta hemoglobin in humans
- Xenology - genes related by gene transfer

Synteny = large regions of chromosomes containing the same genes

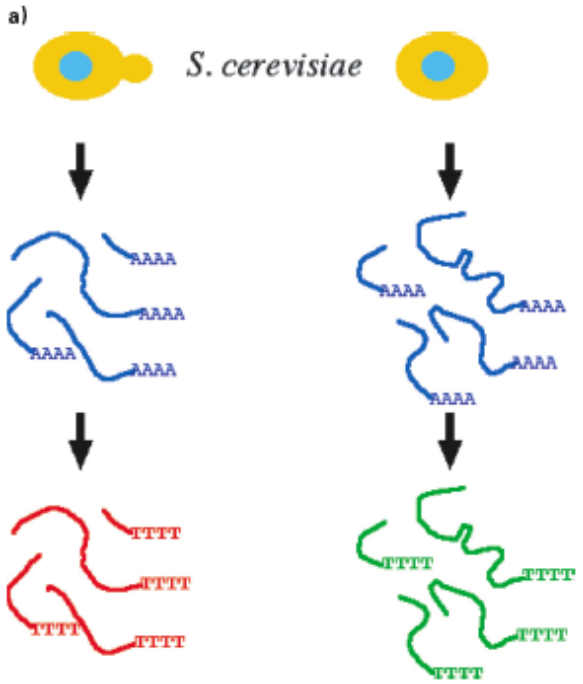


Synteny among Plasmodia

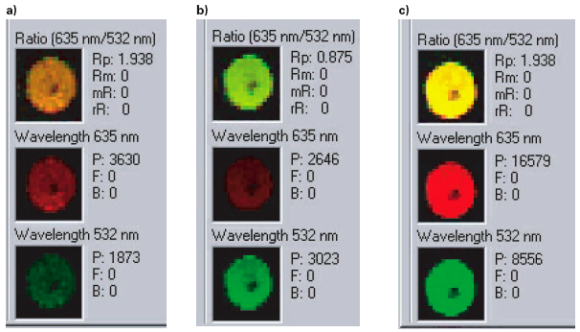


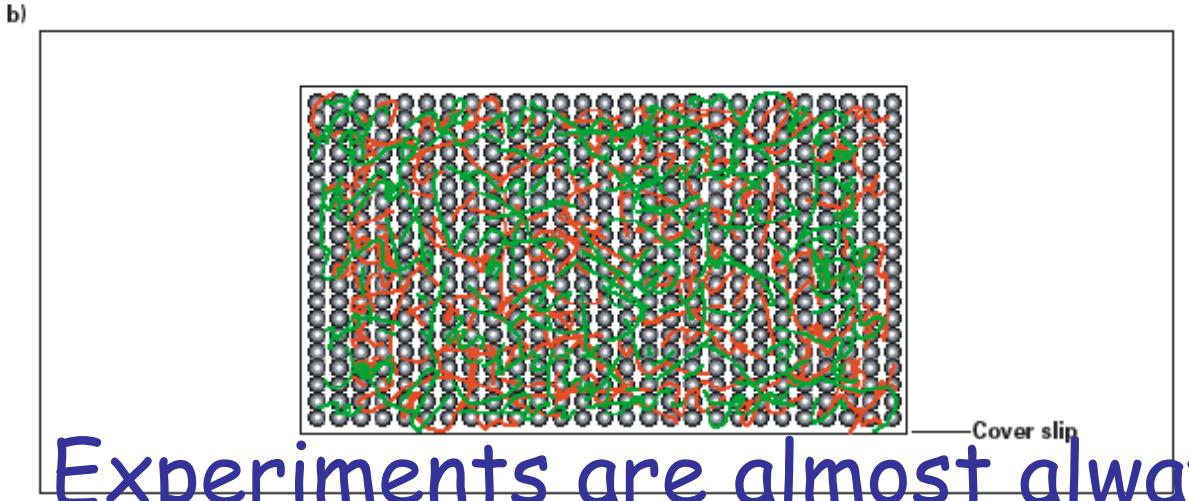
Expression Profiles

- The pattern of expression of one or more genes over time or a set of experimental conditions, e.g. during development or a drug treatment or in a genetic mutant such as a gene knock-out.
- Always… has a time and location component



The RNA samples from the test and the control are labeled with different colors in a reverse-transcription reaction and then hybridized, together, competitively to a slide or chip containing gene sequences in multiple copies.

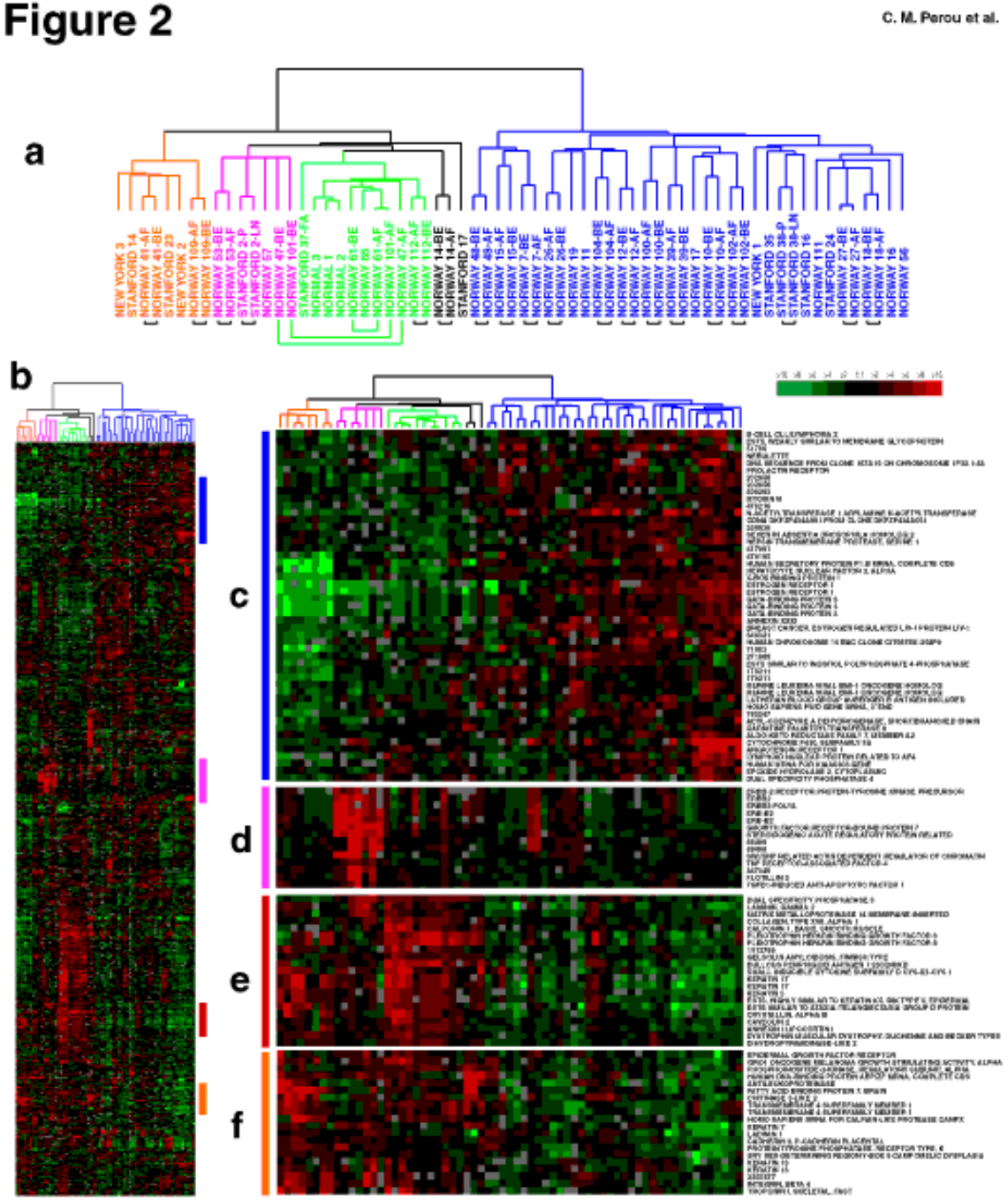




Experiments are almost always Competitions between conditions or stages

Ratios of experimental to control expression are often expressed as colors rather than numbers

Clustered Microarray Data Genes with Similar Expression Profiles are Grouped together



Other RNA expression

- Expressed Sequence Tags, ESTs
  - Usually represent partial cDNA
  - Often clustered
  - Come from libraries that may, or may not be normalized
  - Often used to identify genes in genomes and locations of introns
- SAGE-tags (Serial Analysis of Gene Expression)
  - Primary purpose is relative levels of gene expression
- RNA-Seq (NGS)
  - Little sequence bias
  - Quantitative
  - Can be strand-specific

Genes can be located on either DNA strand

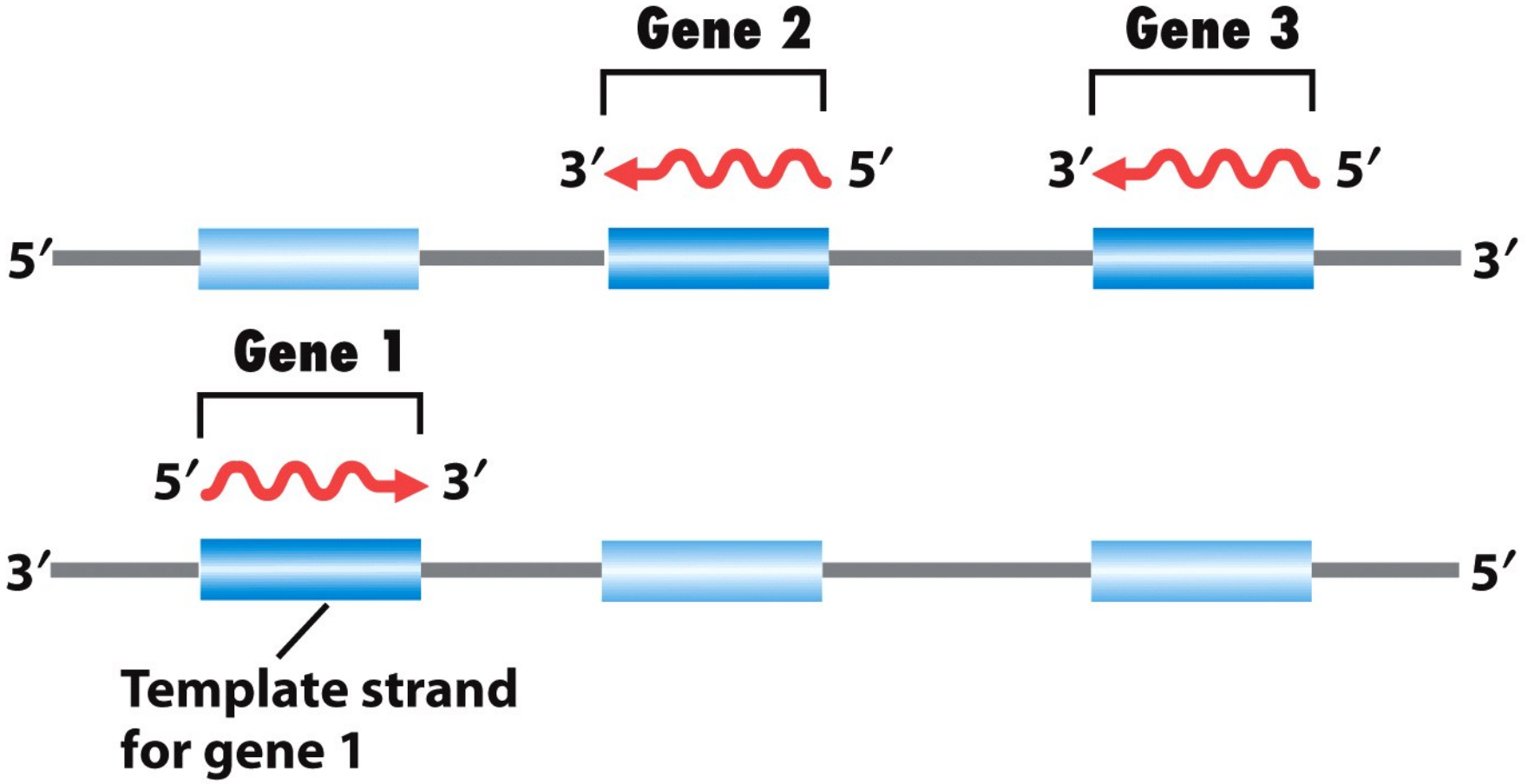


Figure 8-3
Introduction to Genetic Analysis, Ninth Edition

Overview of transcription: Either strand can serve as a template for a gene

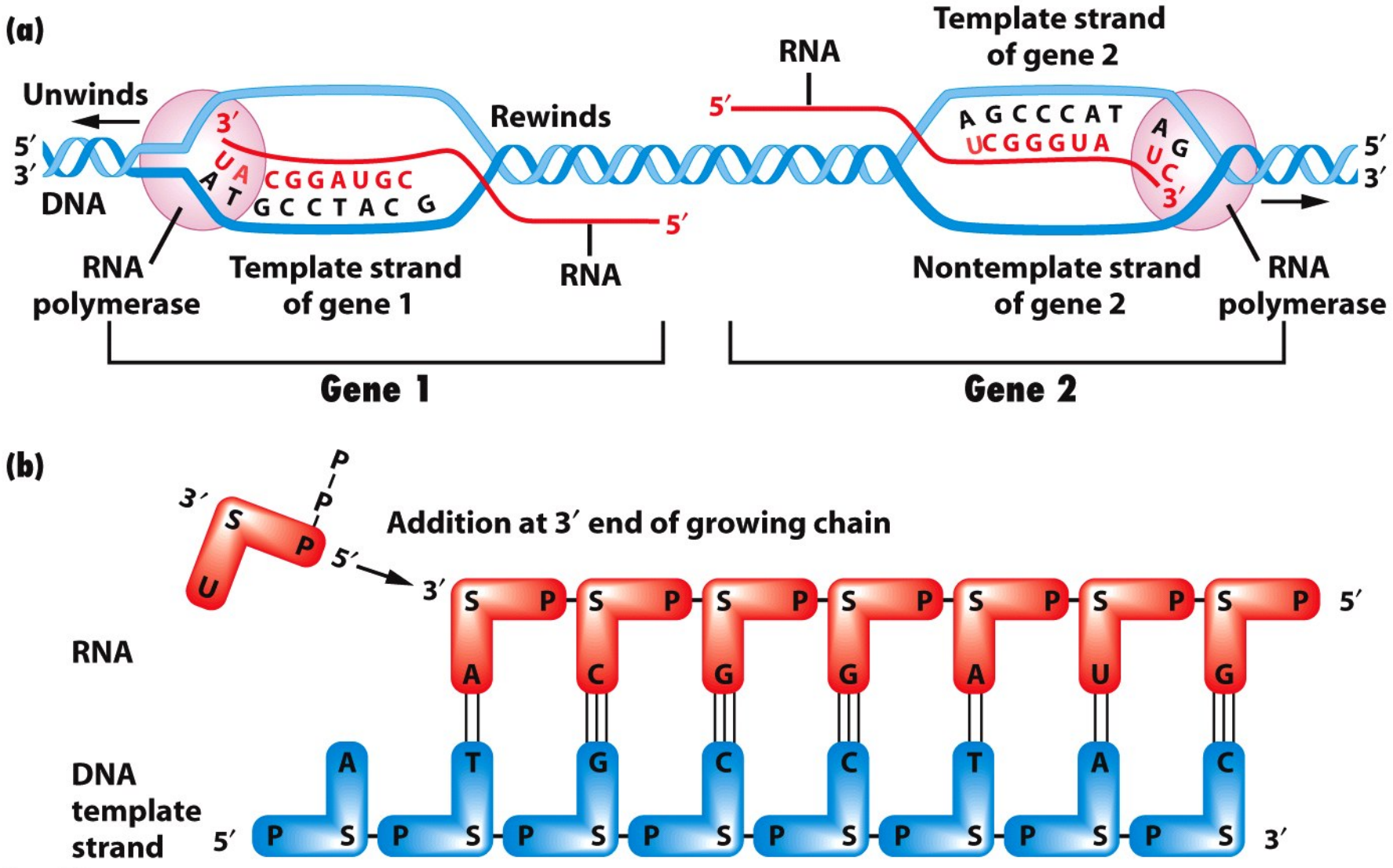


Figure 8-4
Introduction to Genetic Analysis, Ninth Edition

Convention

Gene location = non-template strand, i.e. same as the mRNA

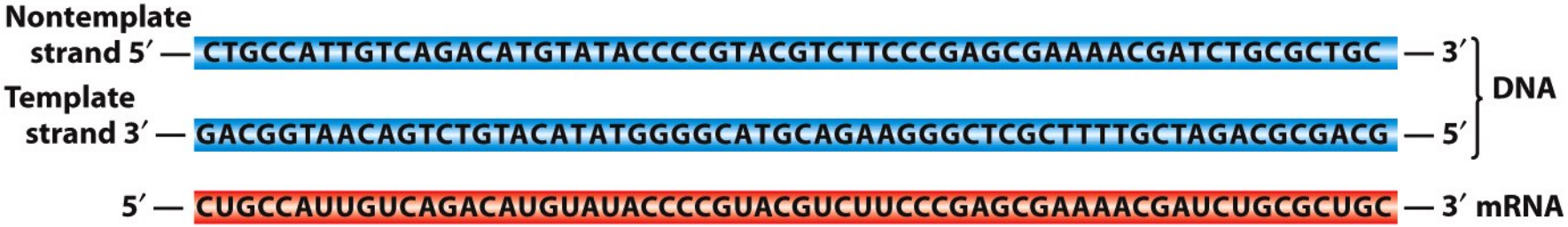


Figure 8-6
Introduction to Genetic Analysis, Ninth Edition

Complex patterns of eukaryotic mRNA splicing: What is a Gene?

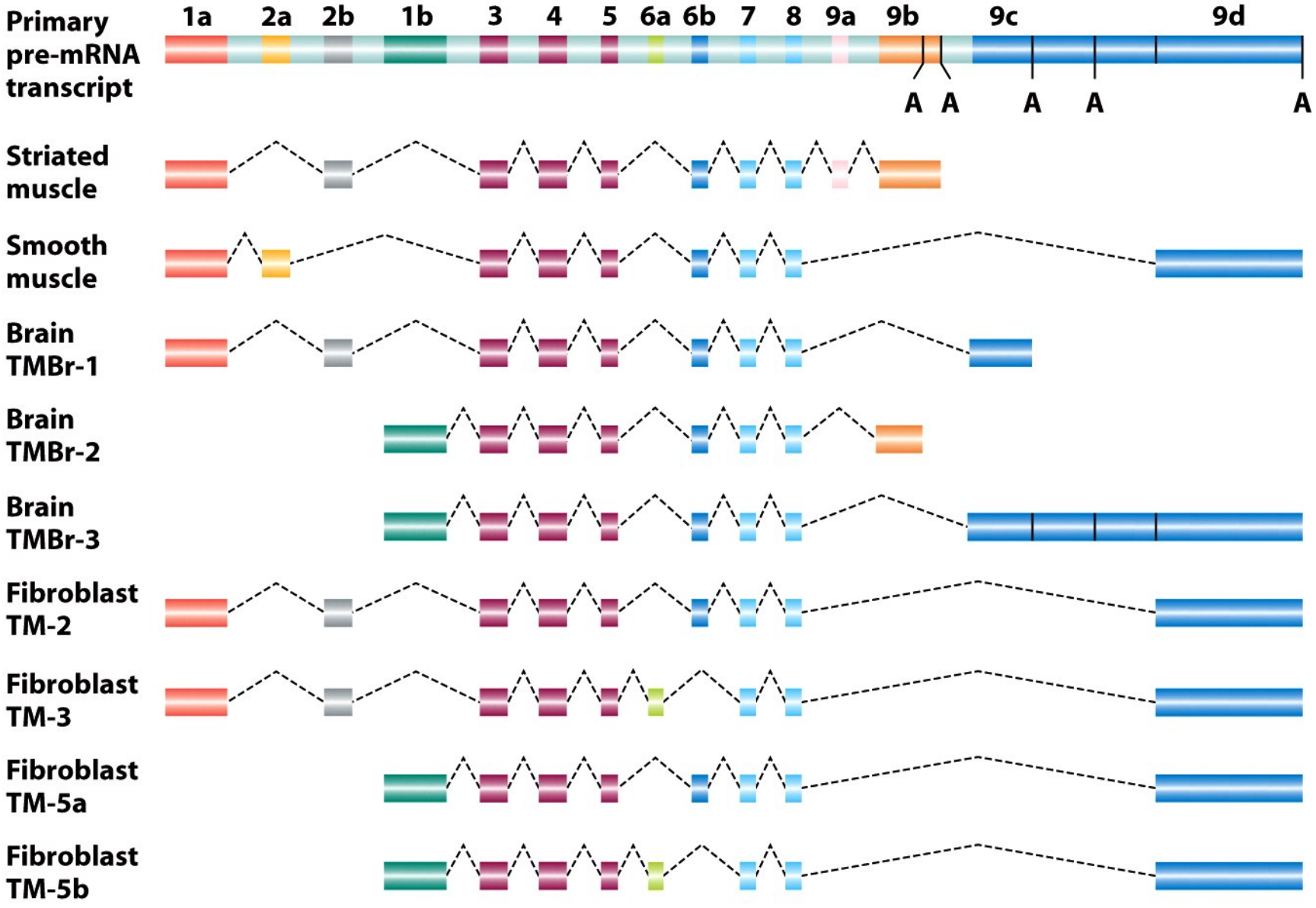


Figure 8-14
Introduction to Genetic Analysis, Ninth Edition

Bioinformatics uses algorithms

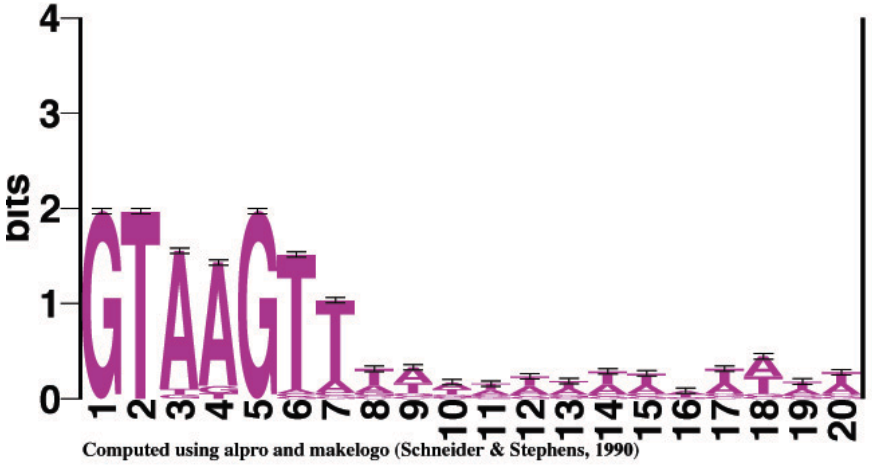
- Algorithms are sets of rules for solving problems or identifying patterns
- Algorithms can be general or case specific and often need to be trained
- Computational analysis, like wet-bench analyses are only as good as the tools, techniques and material allow, and all interpretations come with caveats (like the experimental conditions, often call parameters in bioinformatics.

How to find an intron

- Usually begins with GT and end with AG
- Must be longer than 19 nucleotides
- Must contain a branchpoint "A"
- Donor GT often followed by a sequence pattern. This pattern is species-specific
- Acceptor AG often proceeded by pyrimidine stretch
- Has a mean length of "X" as is observed in this species

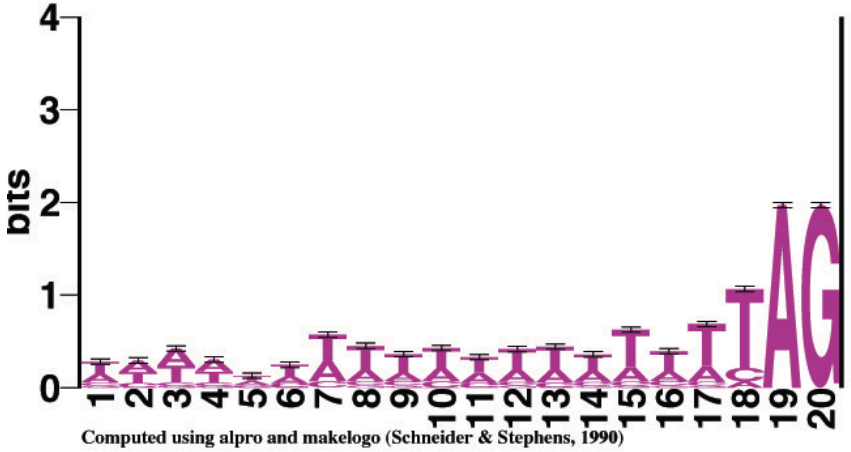
Donor Site

Generated by http://www.bio.cam.ac.uk/seqlogo/log

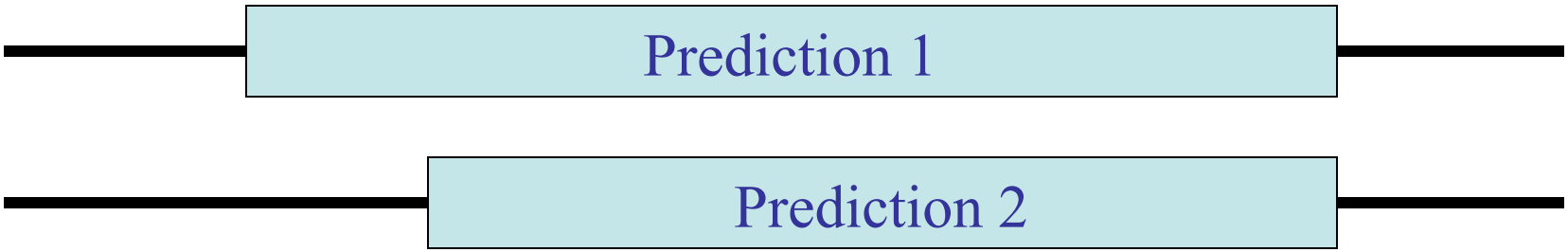


Acceptor site

Generated by http://www.bio.cam.ac.uk/seqlogo/logo.



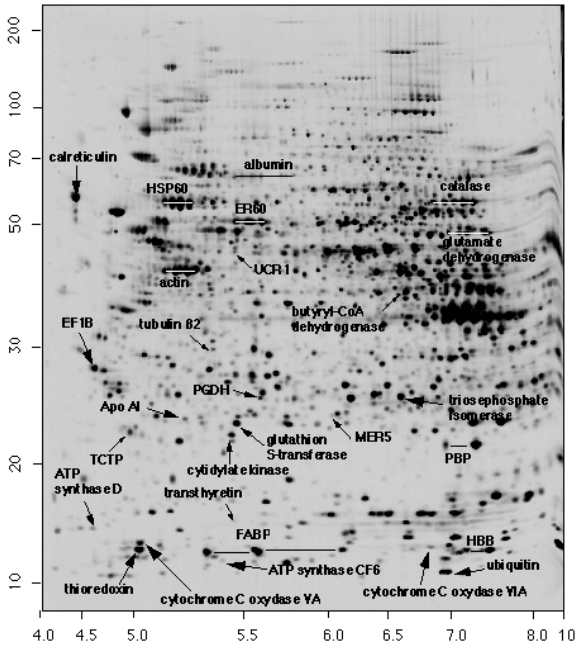
Different prediction methods often generate different results



Protein Expression/Sequence

Data

- MW-Isoelectric point
- MW
- Sequence/spans



Technology

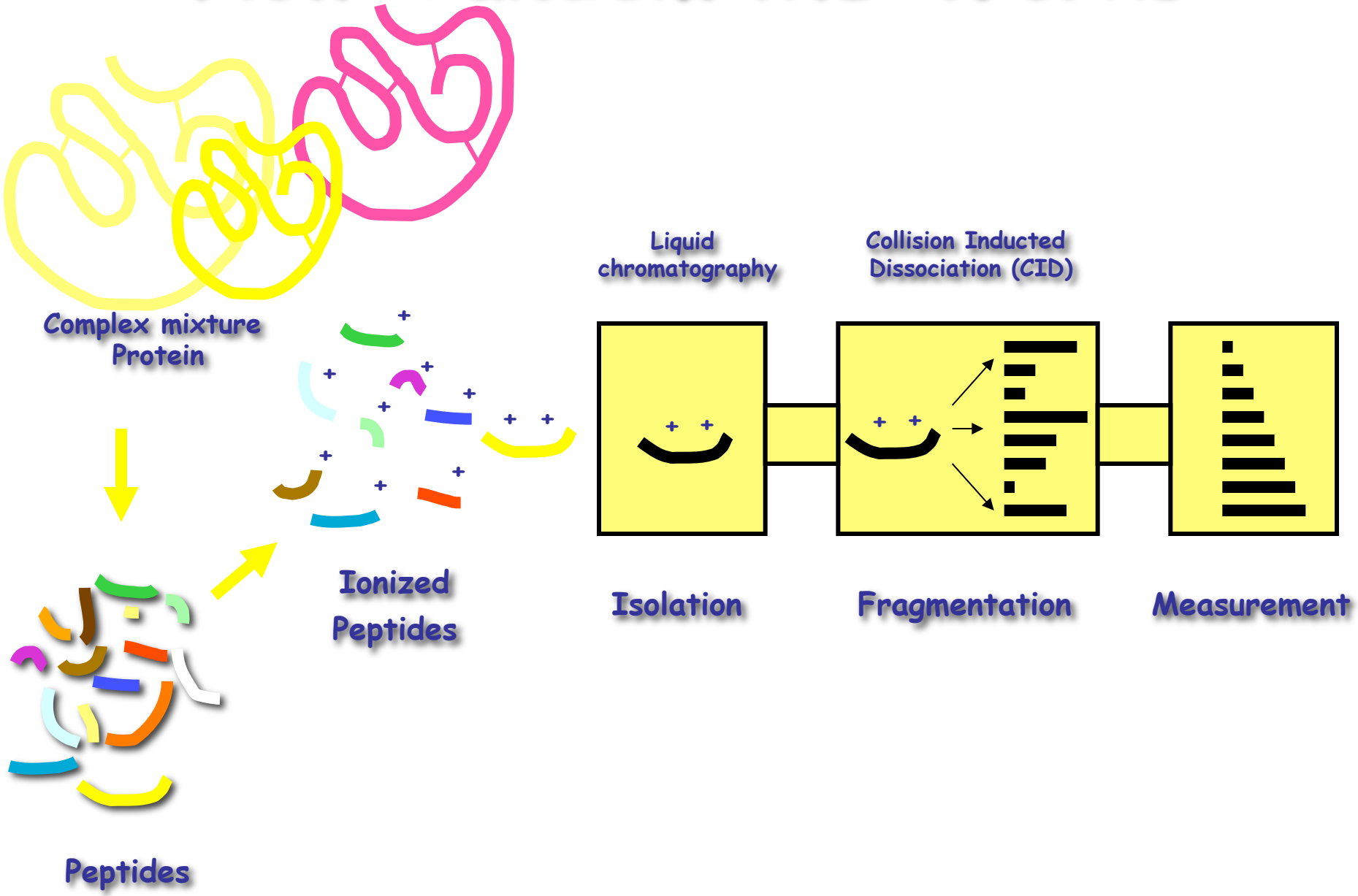
- 2D gel electrophoresis
- Mass spectrometry
- Tandem MS (MS-MS, LC MS-MS etc)

Typical 2 D gel

High throughput mass spectrometry

- Direct identification of proteins from biological sample.
- Capillary liquid chromatography apparatus (LC) coupled with...
- Electrospray tandem mass spectroscopy (MS/MS)
- "Sequest", Mascot, or other software links mass spectra with genomic sequence database.

How Tandem MS Works



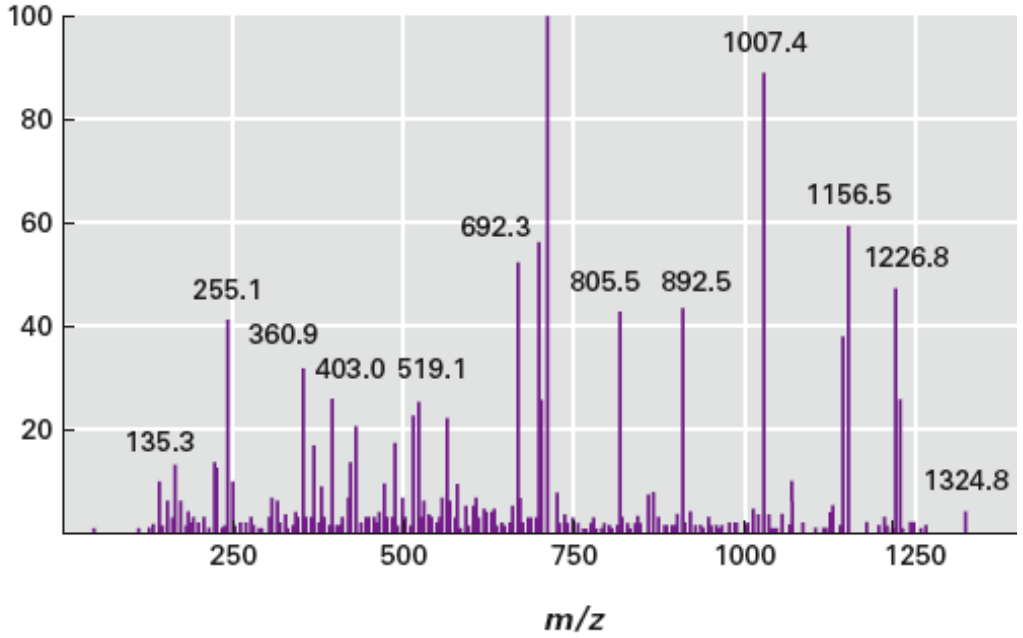
Tandem MS protein data

a)

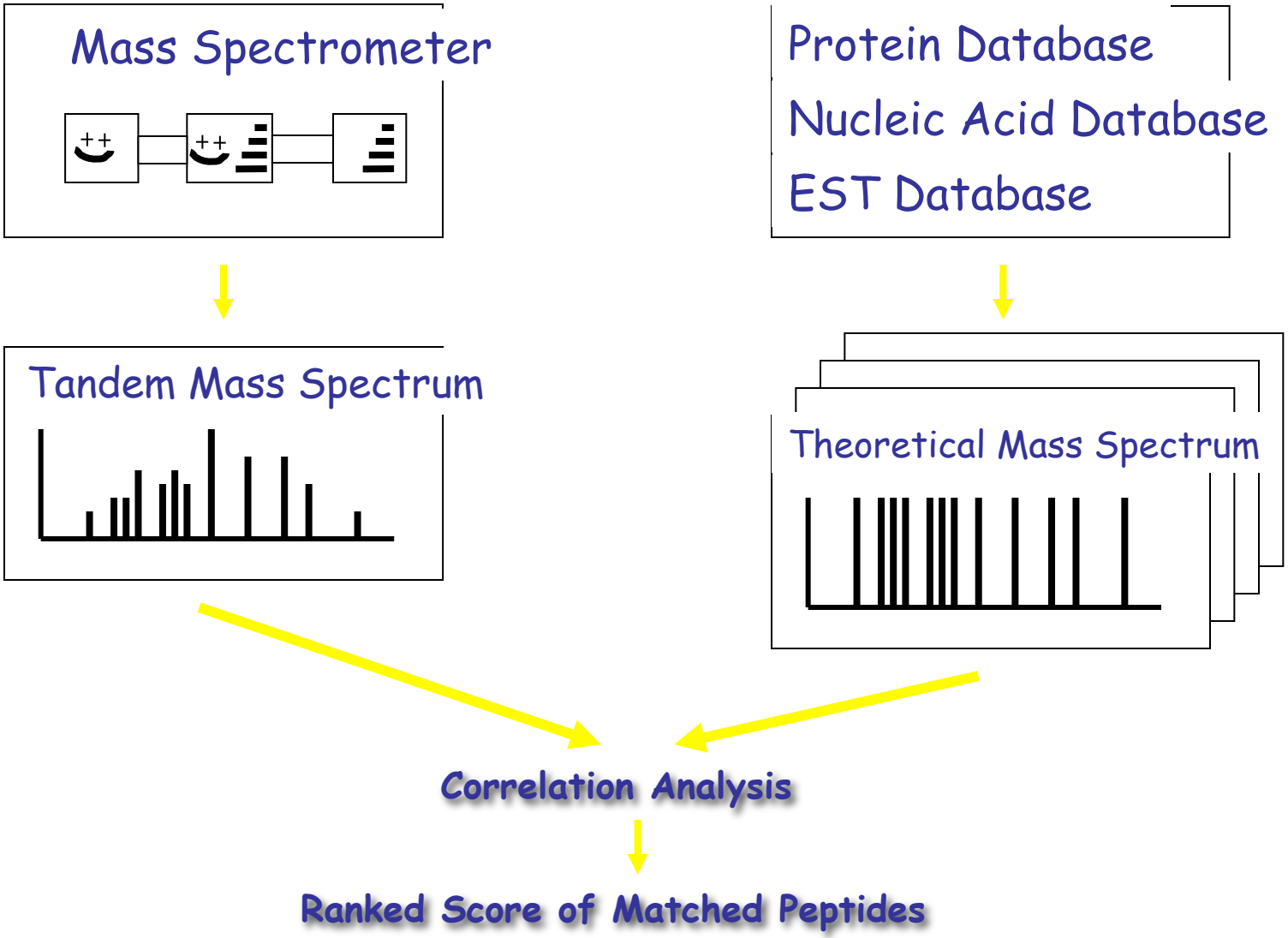
S–P–A–F–D–S–I–M–A–E–T–L–K
(protonated mass 1410.6)

Mass$^+$	b-ions	y-ions	Mass$^+$
81.1	S	PAFDSIMAETLK	1323.6
185.2	SP	AFDSIMAETLK	1226.4
256.3	SPA	FDSIMAETLK	1155.4
403.5	SPAF	DSIMAETLK	1008.2
518.5	SPAFD	SIMAETLK	893.1
605.6	SPAFDS	IMAETLK	806.0
718.8	SPAFDSI	MAETLK	692.3
850.0	SPAFDSIM	AETLK	561.7
921.1	SPAFDSIMA	ETLK	490.6
1050.2	SPAFDSIMAE	TLK	361.5
1151.3	SPAFDSIMAET	LK	260.4
1264.4	SPAFDSIMAETL	K	147.2

b)

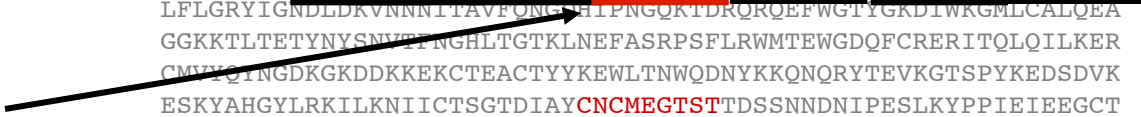


Sequest Database Search



Peptide database

ENNPCKLQYDYNTNVTHGFGQEYPCETDIVERFSDTEGAQCDKKKIKDNSEGACAPYRRL
HVCVRNLENINDYSKINNKHNLLVEVCLAAKYEGESITGRYPQHQETNPDTKSQLCTVLA
RSFADIGDIIRGKDLYRGGNTKEKKKRKKLEENLKTIFGHIYDELKNGKTNGEEELQKRY
RGDKDNDFYQLREDWWDANRETVWKAITCNAGSYQYSQPTCGRGEIPYVTLSKCQCIAGE
VPTYFDYVPQYLRWFEEWAEDFCRKKKKKIPNVKTNCRQVQRGKEKYCDRDGYNCDGTIR
KQYIYRLDTDCTKCSLACKTFAEWIDNQKEQFDKQKQKYQNEISGGGGRRQKRSTHSTKE
YEGYEKHFNEELRNEGKDVRSFLQLLSKEKICKERIQVGEETANYGNFENESNTFSHTEY
CDRCPLCGVDCSSDNCRKKPDKSCDEQITDKEYPPENTTKIPKLTAEKRKTGILKKYEKF
CKNSDGNNGGQIKKWECHYEKNDKDDGNGDINNCIQGDWKTSKNVYYPISYYSFFYGSII
DMLNESIEWRERLKSCINDAKLGKCRKGCKNPCECYKRWVEKKKDEWDKIKEFFRKQKDL
LKDIAGMDAGELLEFYLENIFLEDMKNANGDPKVIEKFKEILGKENEEVQDPLKTKKTID
DFLEKELNEAKNCVEKNPDNECPKQKAPGDGAAPSDPPREDITHHDGEHSSDEDEEEEEE
EEQQPPAEGTEQGEEKSESKEVVEQQETPQKDTEKTVPTTTPTVDVCDTVKTALADTGSL
NAACSLKYVTGKNYGWRCIAPSGTTSGKDGAICVPPRTQELCLYYLKELSDTTQKGLKEA
FIKTAAQETYLLNGKYKEDHQNETASTELLIIDPQTQLNGGGIPEDFKRQMFYTFGDYRD
LFLGRYIGNDLDKVNNNITAVFQNGSHIPNGQKTDRQRQEFWGTYGKDIWKGMLCALQEA
GGKKTLTETYNYSNVTFNGHLTGTKLNEFASRPSFLRWMTEWGDQFCRERITQLQILKER
CMVKQYNGDKGKDDKKEKCTEACTYYKEWLTNWQDNYKKQNQRYTEVKGTSPYKEDSDVK
ESKYAHGYLRKILKNIICTSGTDIAYCNCMEGTSTTDSSNNDNIPESLKYPPIEIEEGCT
CKDPSPGEVIPEKKVPEPKVLPKPPKLPKRQPKERDFPTPALKNAMLSSTIMWSIGIGFA
TFTYFYLKKKTKSTIDLLRVINIPKSDYDIPTKLSPNRYIPYTSGKYRGKRYIYLEGDSG
TDSGYTDHYSDITSSSESEYEELDINDIYAPRAPKYKTLIEVVLEPSGNNTTASGNNTPS
DTQNDIQNDGIPSSKITDNEWNTLKDEFISQYLQSEQPNDVPNDYSSGDIPLNTQPNTLY
FDNPDEKPFITSIHDRDLYSGEEYSYNVNMVNTNNDIPISGKNGTYSGIDLINDSLNSNN

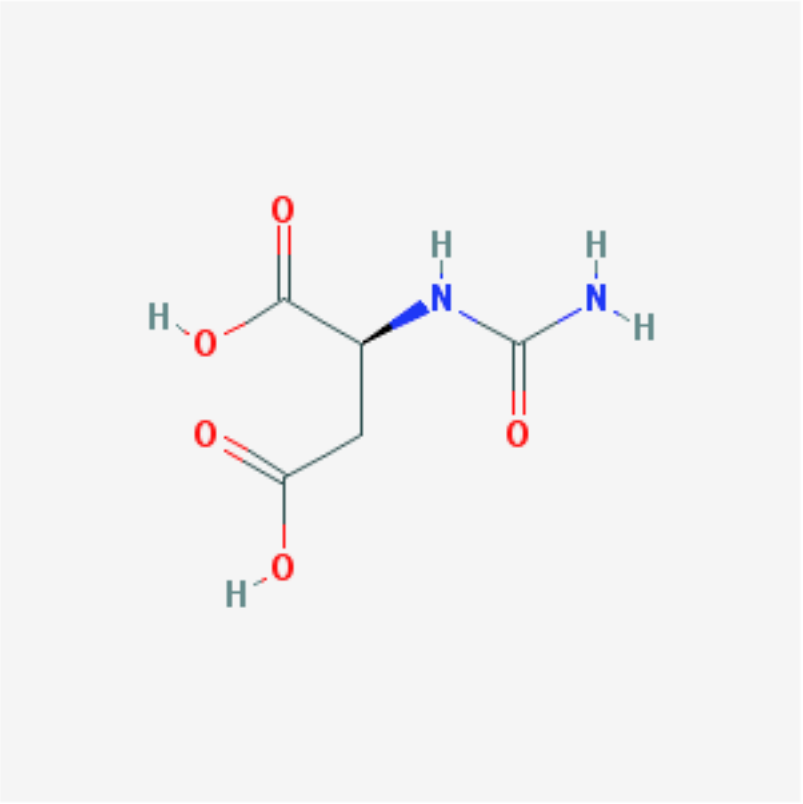


Note: ORFs in addition to predicted Genes must be searched

Overview

PubChem Compound ID:	CID:93072
PubChem Substance ID(s):	3727
Synonyms:	N-Carbamoyl-L-aspartate
Molecular Weight:	176.12742
Molecular Formula:	$C_5H_8N_2O_5$

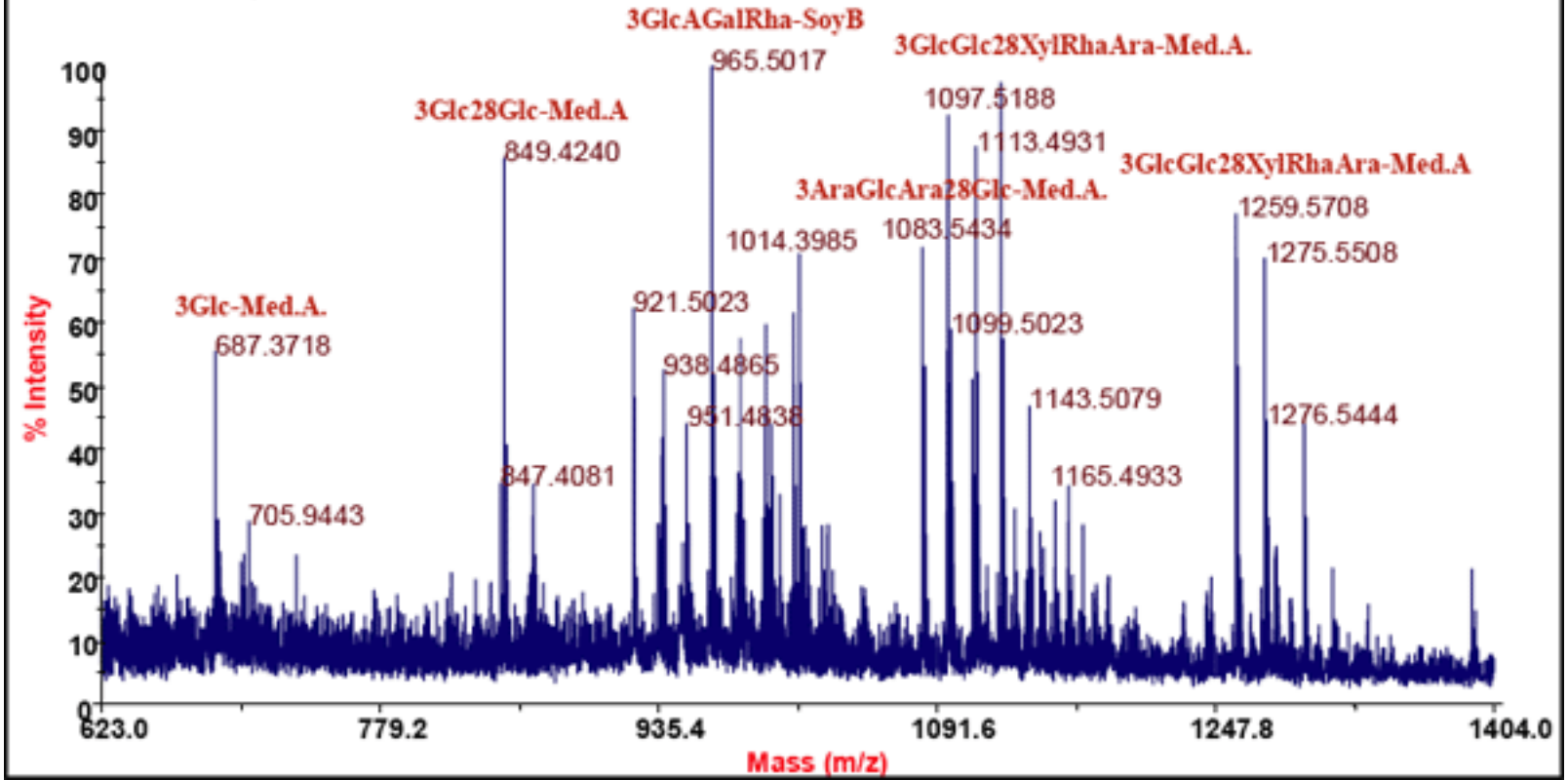
2D Structure



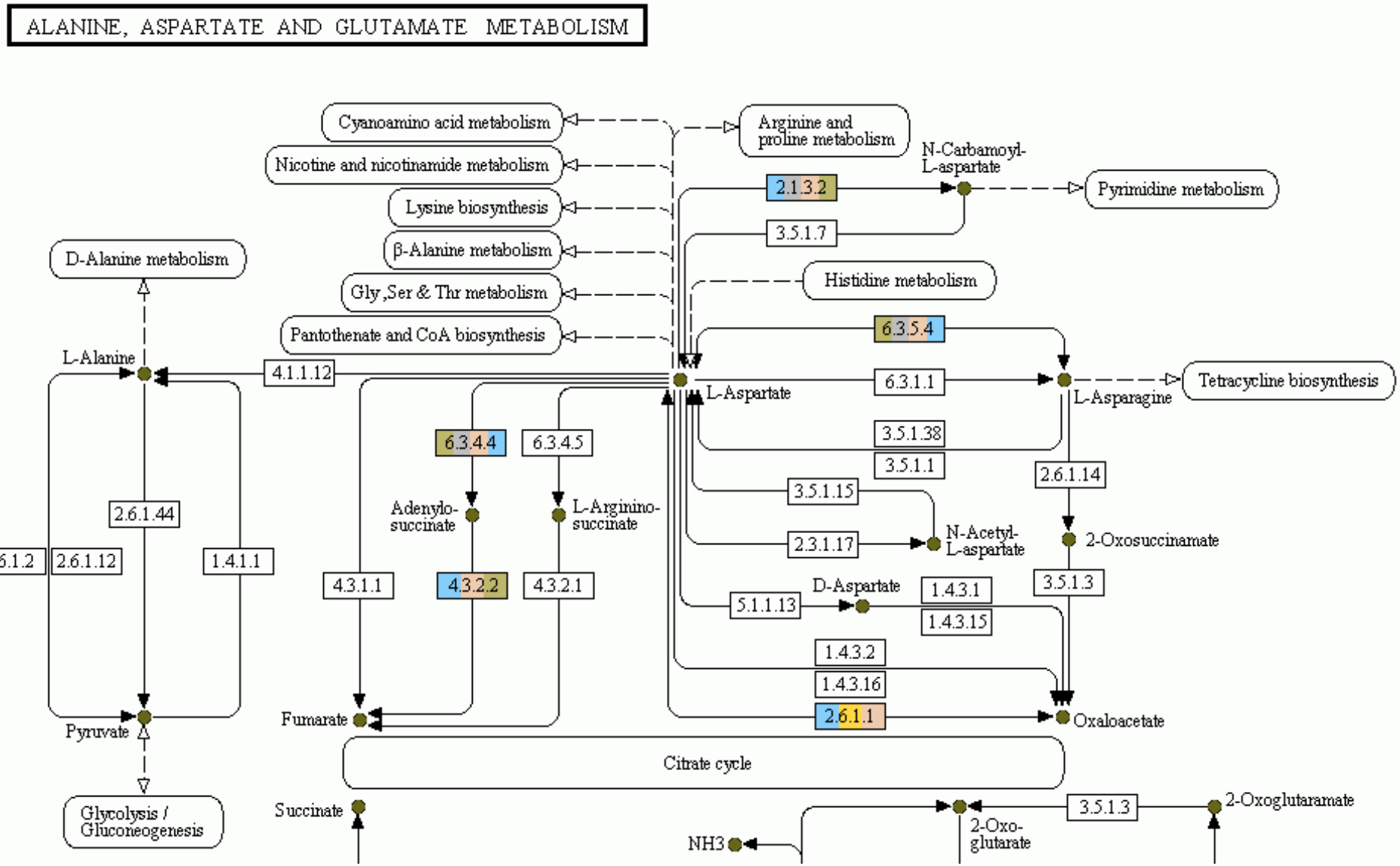
Mass Spectrometry can be used to measure metabolic and other chemical compounds

Complex mixtures can be analyzed and interpreted

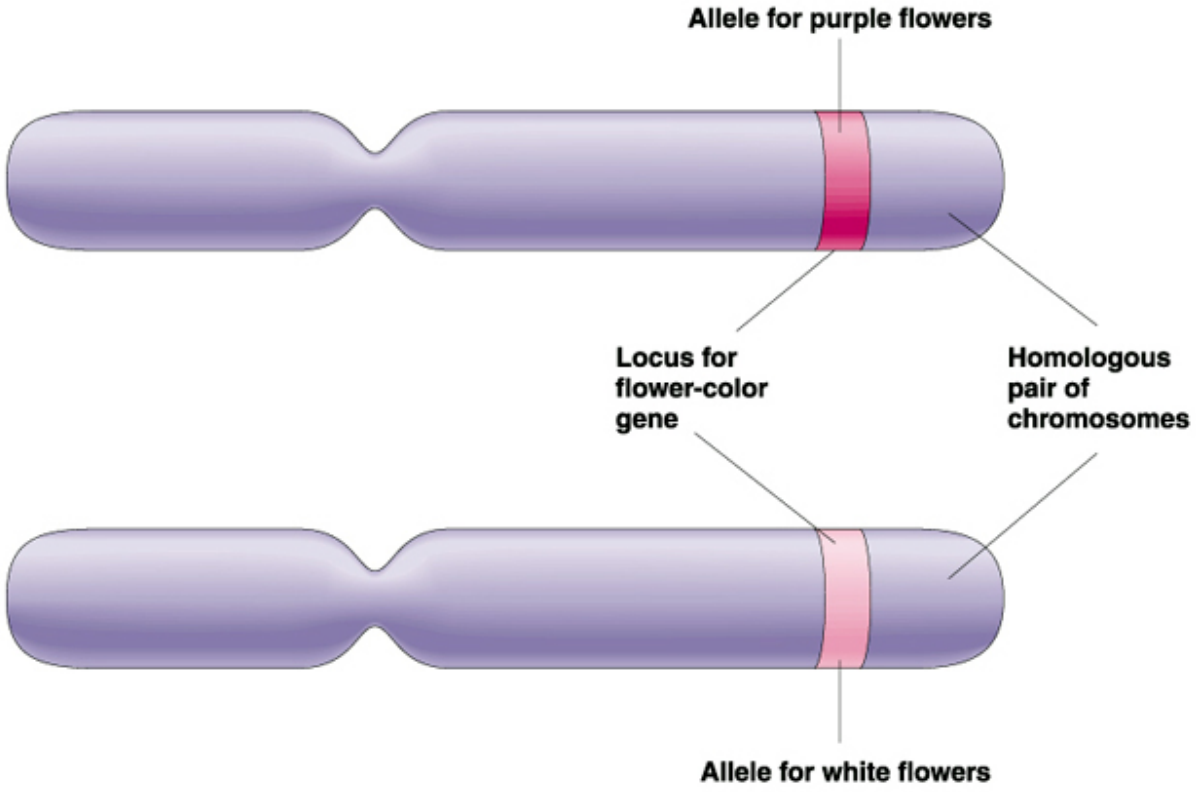
Saponin	Emperical Formula	Theoretical Accurate M/Z	Experimental Accurate M/Z	Error (ppm)
3Glc-Med.A.	C36H56O11Na	687.37201	687.3718	-0.338
3Glc28Glc-Med.A.	C42H66O16Na	849.42483	849.4240	-0.945
3GlcAGalRha-SoyB	C48H78O18Na	965.50856	965.5017	-7.172
3AraGlcAra28Glc-Med.A.	C52H84O22Na	1083.53517	1083.5454	7.715
3GlcGlc28XylRhaAra-Med.A.	C52H82O23Na	1097.51443	1097.5188	3.922
3GlcGlc28XylRhaAra-Med.A	C58H92O28Na	1259.56725	1259.5708	3.126



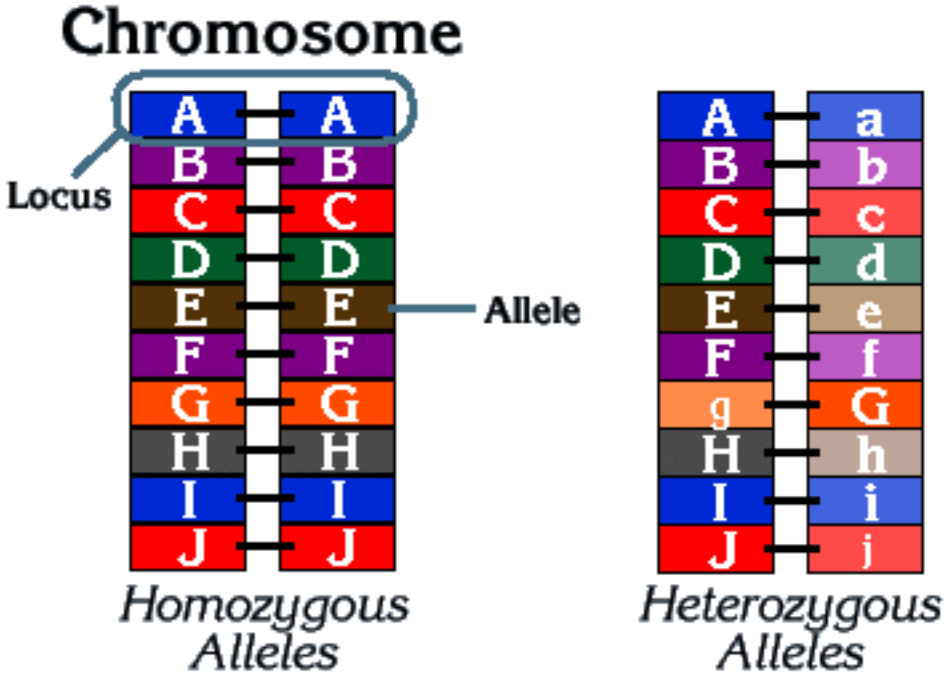
Metabolites can be linked to metabolic pathways and enzymes



Homologous chromsomes (in a diploid)



Loci, alleles and SNPs in a population



A AAGCCTCATC

a ACGCCTCATC

SNP =Single Nucleotide Polymorphism

Alleles and Phenotype

- Some phenotypes are caused by a single locus in the genome and a single allele at that locus (e.g. some flower colors, or *Drosophila* eye color)
- Other phenotypes (Type-I diabetes, heart disease are multi-locus or "complex" (i.e. many genes are involved, each potentially with many alleles)

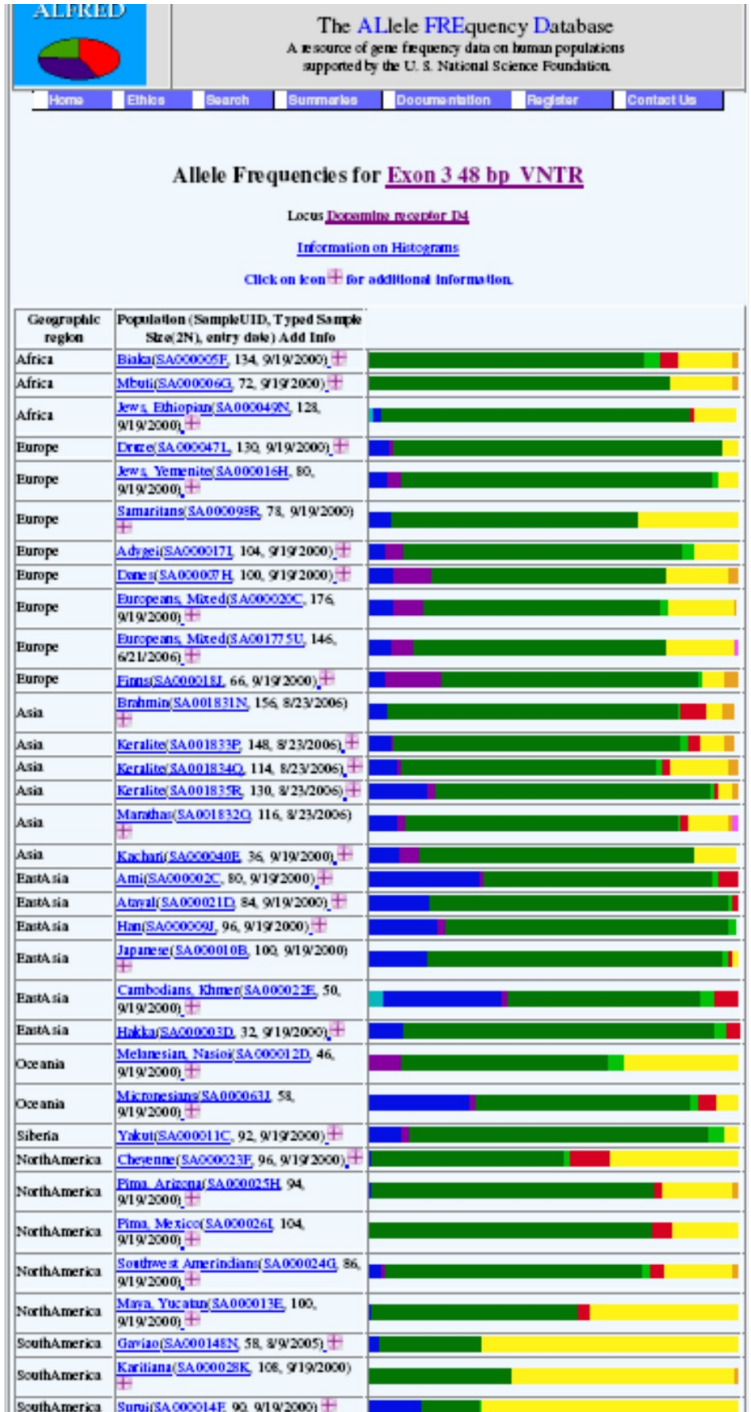
Population data

Data

- Single Nucleotide Polymorphisms, SNPs
- Alleles
- Allele frequency
- Haplotypes

Technology

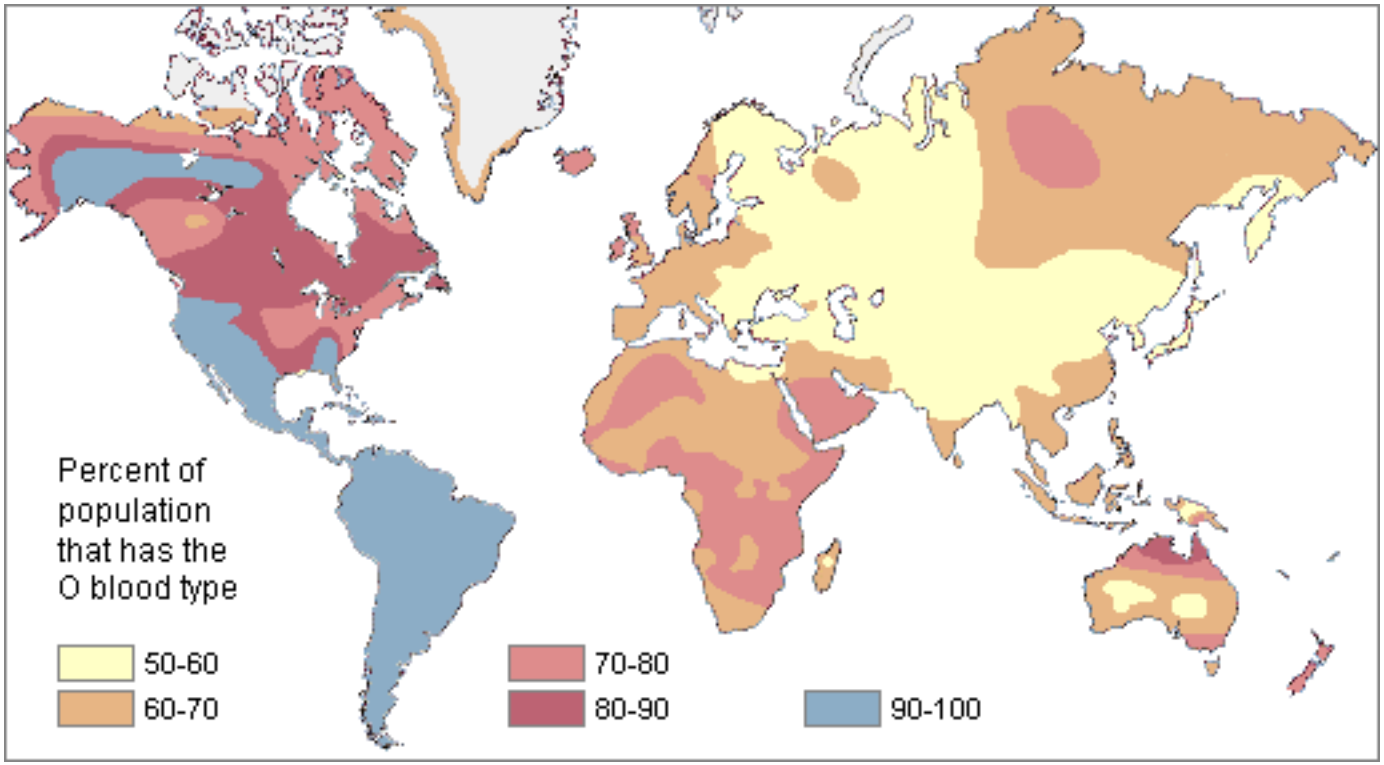
- Chip-Seq
- NGS



Alleles have frequencies in different populations

Populations and alleles have geographic boundaries

A parasite isolate comes from a particular population, a particular location and will have a specific haplotype (e.g. representation of alleles) often characterized via SNPs



Parasite Isolates

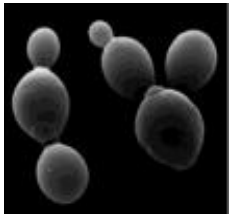
Data

- Species, Strain,
- Isolate
- Location, Date
- SNP
- Sequence
- Allele
- phenotpe

Technology

- PCR-RFLP
- Microsatellites
- Sequencing
- SNP chip
- GPS

Infectious Disease Paradigm



Experimental systems

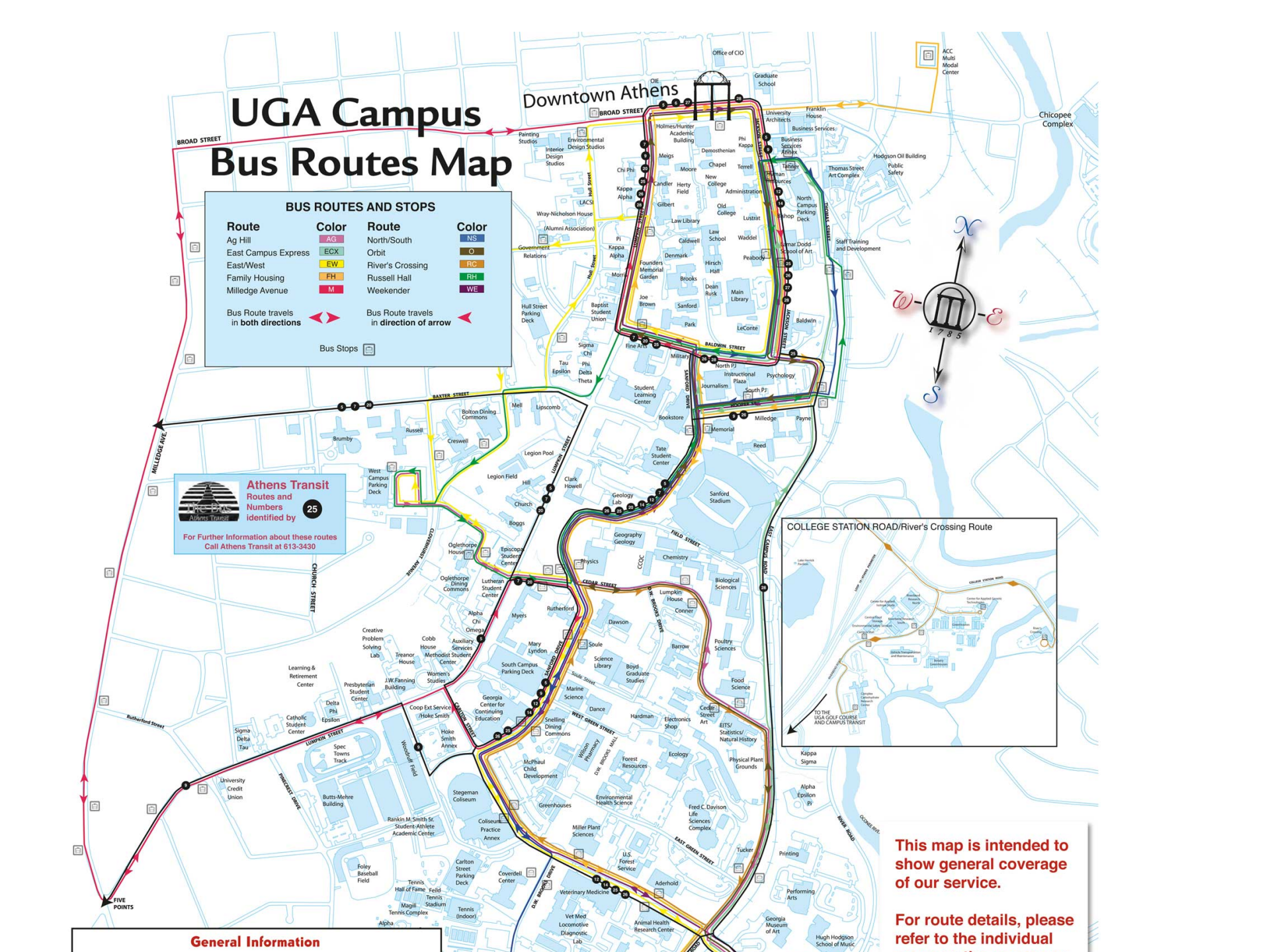
Host

Pathogen

Vector

Important Information

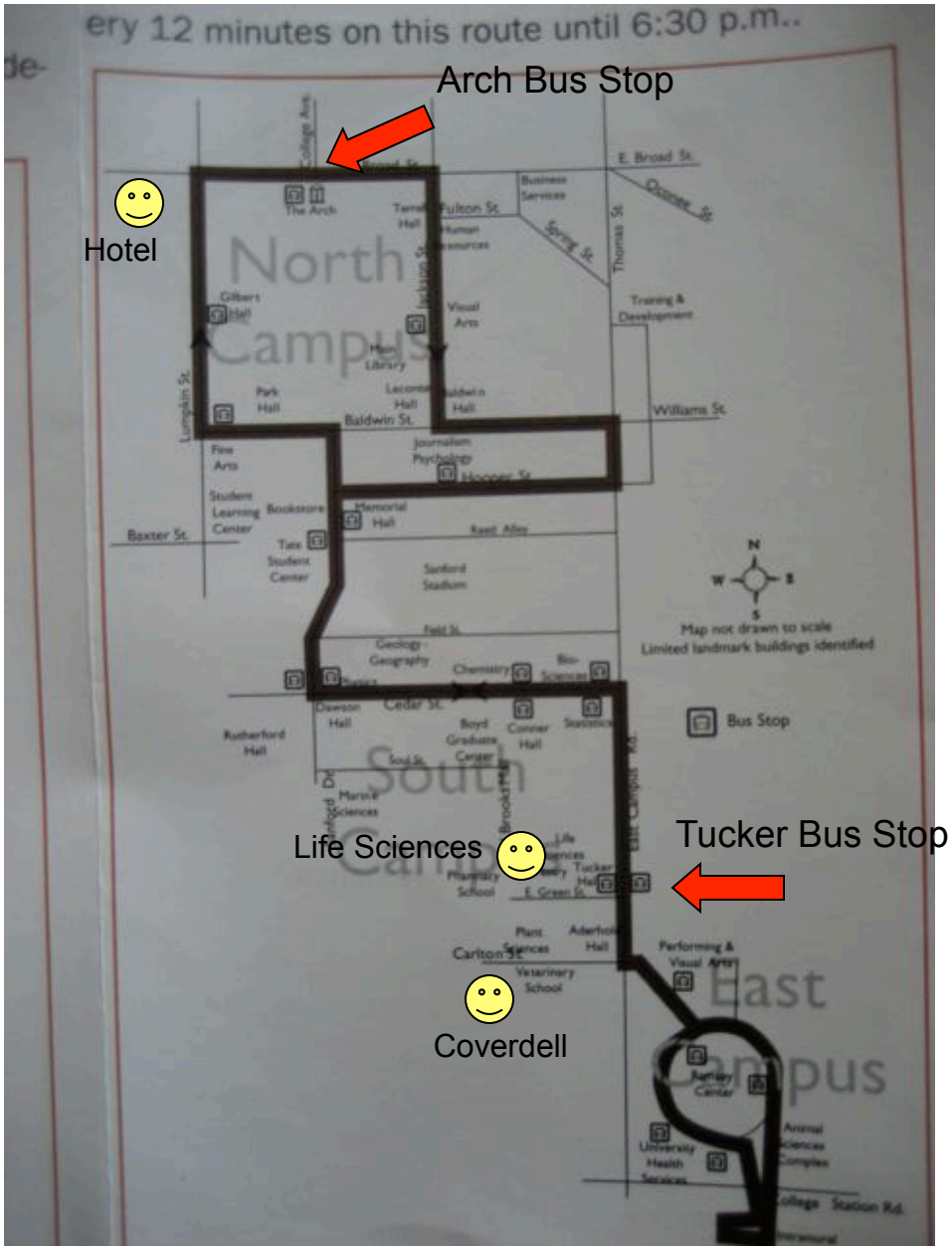
- It might rain, especially Monday AM
- Air conditioned buildings may be cold, you may want a sweater
- You do not need your laptop for workshop sessions
- Buses are infrequent because we are in "Intersession"
- Guides (Brian and Omar) will leave hotel tomorrow lead folks to the workshop room via University bus


UGA Campus
Bus Routes Map
Downtown Athens
BUS ROUTES AND STOPS
Route
Color
Ag Hill
AG
East Campus Express
ECX
East/West
EW
Family Housing
FH
Milledge Avenue
M
North/South
NS
Orbit
O
River's Crossing
RC
Russell Hall
RH
Weekender
WE
Bus Route travels in both directions
Bus Route travels in direction of arrow
Bus Stops
Athens Transit
Routes and Numbers identified by
25
For Further Information about these routes Call Athens Transit at 613-3430
COLLEGE STATION ROAD/River's Crossing Route
TO THE UGA GOLF COURSE AND CAMPUS TRANSIT
This map is intended to show general coverage of our service.
For route details, please refer to the individual
General Information
BROAD STREET
BAXTER STREET
BALDWIN STREET
FIELD STREET
CEDAR STREET
WEST GREEN STREET
EAST GREEN STREET
LUMPKIN STREET
CHURCH STREET
MILLEDGE AVE
RUTHERFORD STREET
FIVE POINTS
Office of CIO
OIE
Graduate School
ACC Multi Modal Center
Chicopee Complex
Franklin House
University Architects
Business Services
Business Services Annex
Hodgson Oil Building
Public Safety
Thomas Street Art Complex
Staff Training and Development
Painting Studios
Interior Design Studios
Environmental Design Studios
Holmes/Hunter Academic Building
Phi Kappa
Demosthenian
Chapel
Terrell
Meigs
Moore
New College
Herty Field
Chi Phi
Kappa Alpha
Candler
Administration
Gilbert
Old College
Lustrat
Thomas
Human Resources
North Campus Parking Deck
Bishop
Wray-Nicholson House
(Alumni Association)
Government Relations
LACS
Hull Street
Law Library
Law School
Caldwell
Waddel
Pi Kappa Alpha
Denmark
Peabody
Lumpkin Street
Jackson Street
Founders Memorial Garden
Morris
Brooks
Hirsch Hall
Dean Rusk
Main Library
Lamar Dodd School of Art
Hull Street Parking Deck
Baptist Student Union
Joe Brown
Sanford
Park
LeConte
Baldwin
Sigma Chi
Tau Epsilon
Phi Delta Theta
Fine Arts
Military
North PJ
Instructional Plaza
Psychology
Journalism
South PJ
Sanford Drive
Student Learning Center
Bookstore
Milledge
Payne
Memorial
Reed
Tate Student Center
Bolton Dining Commons
Mell
Lipscomb
Russell
Brumby
Creswell
Legion Pool
Legion Field
Hill
Church
Boggs
Clark Howell
West Campus Parking Deck
Geology Lab
Geology
Geography
Physics
Chemistry
CCQC
Sanford Stadium
Oglethorpe House
Episcopal Student Center
Oglethorpe Dining Commons
Lutheran Student Center
Cloverhurst Avenue
Rutherford
Myers
Alpha Chi Omega
Creative Problem Solving Lab
Cobb House
Auxiliary Services
Methodist Student Center
Treanor House
Mary Lyndon
Soule
Dawson
Science Library
Boyd Graduate Studies
Lumpkin House
Conner
Barrow
Poultry Sciences
Biological Sciences
Food Science
Learning & Retirement Center
Presbyterian Student Center
J.W. Fanning Building
Women's Studies
South Campus Parking Deck
Soule Street
Marine Science
Georgia Center for Continuing Education
Coop Ext Service /Hoke Smith
Hoke Smith Annex
Delta Phi Epsilon
Catholic Student Center
Sigma Delta Tau
Spec Towns Track
Woodruff Field
Snelling Dining Commons
Dance
Hardman
Electronics Shop
Cedar Street Art
EITS/ Statistics/ Natural History
Physical Plant Grounds
McPhaul Child Development
Wilson Pharmacy
D.W. Brooks Mall
Forest Resources
Ecology
Environmental Health Science
Greenhouses
Miller Plant Sciences
Fred C. Davison Life Sciences Complex
Tucker
Printing
Performing Arts
Kappa Sigma
Alpha Epsilon Pi
River Road
University Credit Union
Butts-Mehre Building
Rankin M. Smith Sr. Student-Athlete Academic Center
Coliseum Practice Annex
Stegeman Coliseum
Pinecrest Drive
Foley Baseball Field
Tennis Hall of Fame
Magill Tennis Complex
Tennis Field
Tennis Stadium
Tennis (Indoor)
Carlton Street Parking Deck
Coverdell Center
U.S. Forest Service
Veterinary Medicine
Aderhold
Vet Med Locomotive Diagnostic Lab
Animal Health Research Center
Georgia Museum of Art
Hugh Hodgson School of Music
Alpha
N
W
E
S
1785

Orbit Bus Route
NOT Running

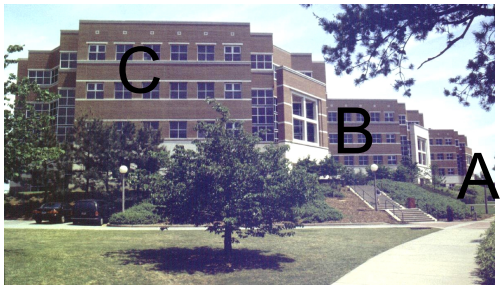
3 blocks, .6 Km

~10 blocks
1.1 miles



Get off at Tucker Hall

Return to Hotel



Life Sciences

Coverdell

Take the Orbit bus from the Arch to Tucker Hall at East Campus and Green Street
The Life Sciences building is on Greene street, up the hill on the Right Rm C128

General Schedule

- Workshop all day Mon – Wed Schedule in your packets
- Dinner at the Coverdell Center on Monday (Informal!)
- Breakfast (2 choices)
  - Holiday Inn (coupons received at check-in
  - Have nibbles with morning coffee break
- Coffee and lunch provided (M-W). Tuesday and Wednesday Dinner is on your own.

Questions?

Now that you know us, We would like to know you

- Please tell us your name and a bit about your research.

- Also, if there is something specific that you came here to learn, please state it so that we can cover it if possible