BioPerl: Pairwise Sequence Alignment

```
use Bio::Tools::pSW;

$factory = new Bio::Tools::pSW( '-matrix' =>
   'blosum62.bla', '-gap' => 12, '-ext' => 2, );

$factory->align_and_show($seq1, $seq2, STDOUT);
```

BioPerl: Running BLAST

```
# This program only shows how to invoke BLAST and store the result
use Bio::SeqIO;
use Bio::Tools::Run::RemoteBlast;
my $Seq_in = Bio::SeqIO->new (-file => $ARGV[0], -format => 'fasta');
my $query = $Seq_in->next_seq();
my $factory = Bio::Tools::Run::RemoteBlast->new( '-prog' => 'blastp',
    '-data' => 'swissprot', _READMETHOD => "Blast" );
my $blast_report = $factory->submit_blast($query);
my $result = $blast_report->next_result;
while( my $hit = $result->next_hit()) {
    print "\thit name: ",
    $hit->name(), " significance: ", $hit->significance(), "\n";
}
# There are programs on the bioperl website that can help you automatically
    parse the information returned by BLAST.
```

BioPerl: Multiple Sequence Alignment

```
@params = ('ktuple' => 2, 'matrix' => 'BLOSUM');
$factory =
Bio::Tools::Run::Alignment::Clustalw->new(@params);
$aln = $factory->align(\@seq_array);


foreach $seq ( $aln->eachSeq() ) {
        print $seq->seq(), "\n"; }
```

BioPerl: Structure

- Ability to store and manipulate structures.
- Modules: Atom, Chain, Residue, Model, Entry, IO
- Atom
  - new, x, y, z, xyz, residue, element,
- Chain, Residue
- Entry
  - Add_model, chain, add_chain, residue, add_residue, get_residue, add_atom, get_atoms, conect, get_atom_by_serial, seqres, …
- Model

BioPerl: Structure

```
use Bio::Structure::IO;
$in = Bio::Structure::IO->new(-file => "inputfilename" , '-format' => 'pdb');
$out = Bio::Structure::IO->new(-file => ">outputfilename" , '-format' => 'pdb');
# note: we quote -format to keep older perl's from complaining.
while ( my $struc = $in->next_structure() ) {
    $out->write_structure($struc);
    print "Structure ",$struc->id," number of models: ",
            scalar $struc->model,"\n";
}
```

More Bioperl Modules

Bioperl-1.0.2::Bio::Structure::SecStr::DSSP
bioperl-1.0.2::Bio::Structure::SecStr::STRIDE
bioperl-1.0.2::Bio::Symbol
bioperl-1.0.2::Bio::Tools
bioperl-1.0.2::Bio::Tools::Alignment
bioperl-1.0.2::Bio::Tools::Bplite
bioperl-1.0.2::Bio::Tools::Blast
bioperl-1.0.2::Bio::Tools::HMMER
bioperl-1.0.2::Bio::Tools::Prediction
bioperl-1.0.2::Bio::Tools::Run::Alignment
bioperl-1.0.2::Bio::Tools::Sim4
bioperl-1.0.2::Bio::Tools::StateMachine
bioperl-1.0.2::Bio::Tree
bioperl-1.0.2::Bio::TreeIO

Cystic Fibrosis

- Before 1985
  - Genetic Disorder that makes people susceptible to chronic pulmonary infection
  - 1/25 Caucasians carriers, 1/2500 suffer from CF
- 1985
  - 3 groups independently showed it is on 7^{th} chromosome
- 1989
  - 1480 aa, 27 exons, 23 Kb gene
- Since 1989
  - Many mutations and their effects
  - Gene Therapy

Types of DNA sequence data

- EST/STS
- Single gene sequences
- Pre-finished DNA clone sequences
- Large genomic sequence tracts

- Ordering information missing in all except last two above.

Types of Maps

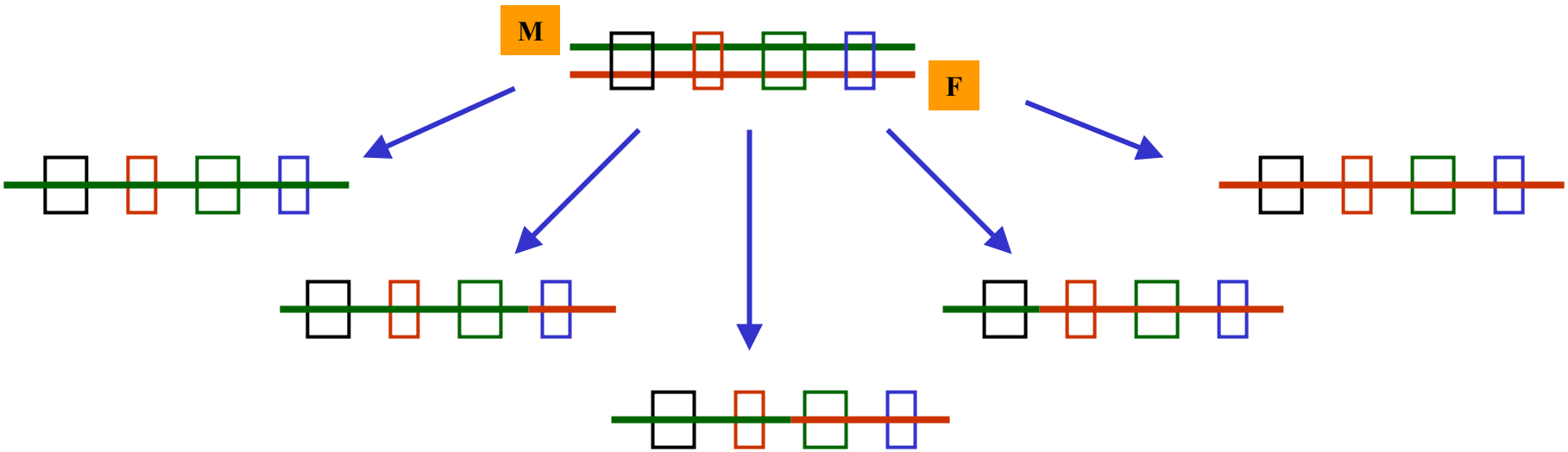
- Cytogenetic Maps
  - Labeled probe hybridized to chromosome, thus assigning probe to region of chromosome
- Genetic Linkage Maps
  - Recombination crossover events in multi-generational families used to infer relative order of markers
- Radiation Hybrid Maps
  - Radiation used to break chromosomes
  - PCR strength patterns used to determine proximity

Types of Maps

- EST Map
- Physical Maps
  - Precise location of various markers & genes
- Comparative Maps
  - Comparing genomes of organisms
- Integrated Maps
  - Connecting & combining different maps

Genetic Mapping

- To "approximately" locate gene in genome
- To find the order of the genes in genome
  - Sturtevant, 1913: First genetic map for 6 genes in fruit fly.



- Higher the frequency of "coinheritance" of alleles of two genes, the closer they are likely to be.

Genetic Mapping

- Analyze patterns of disease inheritance in several generations of an affected family.
- Determine which genes have alleles that are frequently co-inherited with the disease.
- This will help pin down relative location of gene on genetic map.
- Find EST that exhibits systematic difference between diseased and normal individuals and then locate gene around EST.

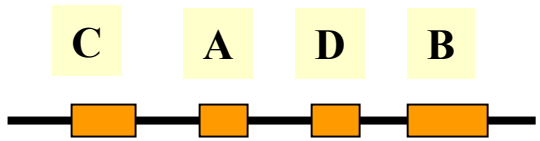
Why genetic & physical maps?

- Determine which genomic elements present in a genomic region.
- Determine their relative order.
- Determine chromosomal positions of genes.

Physical Mapping

- Determine **exact** physical location of landmarks/markers/features of interest.
  - STS, EST, Gene, etc.

		STS			
		A	**B**	**C**	**D**
Clones Fragments	**1**	**1**	**0**	**1**	**0**
	2	**1**	**1**	**0**	**1**
	3	**1**	**0**	**0**	**1**



Hybridization Method

	A	B	C	D	E	F	G
1	1		1				
2	1	1			1		1
3		1	1	1	1	1	1
4		1	1	1		1	1
5			1	1		1	1
6	1			1		1	1
7	1	1		1	1		1
8	1	1		1			1
9				1			

ACABEGBCDEFGBCDFG…

CAEBGCFDAGEBAGD

{CA}
{C}{A}{BEG}
CA{BEG}{CDF}
CAE{BG}{CDF}{AG}
CAE{BG}C{DF}{AG}{BE}
CAE{BG}CFD{AG}BE

Partial Digest Problem

- A restriction enzyme cuts at restriction sites.
- It produces fragments of specific lengths.

X = {7, 47, 137, 177}

ΔX = {7, 40, 47, 90, 100, 130, 137, 140, 170, 177, 230, 270, 277}

Partial Digest Problem: Given ΔX , find X.

Double Digest Problem

- Input: 2 Restriction Enzymes A & B

  Fragments(A)

  Fragments(B)

  Fragments(A & B)

- Output: Reconstruction
  - NP-Complete [Goldstein & Waterman]
  - Interesting Transformation [Schmitt & Waterman]
  - Generating all solutions [Pevzner]

Errors in practice in Physical Mapping

- False Positives
- False Negatives
  - Improper hybridization conditions
- Chimeric Clones
  - 2 clones may join together even though they are not “adjacent”
- Contamination from Host DNA