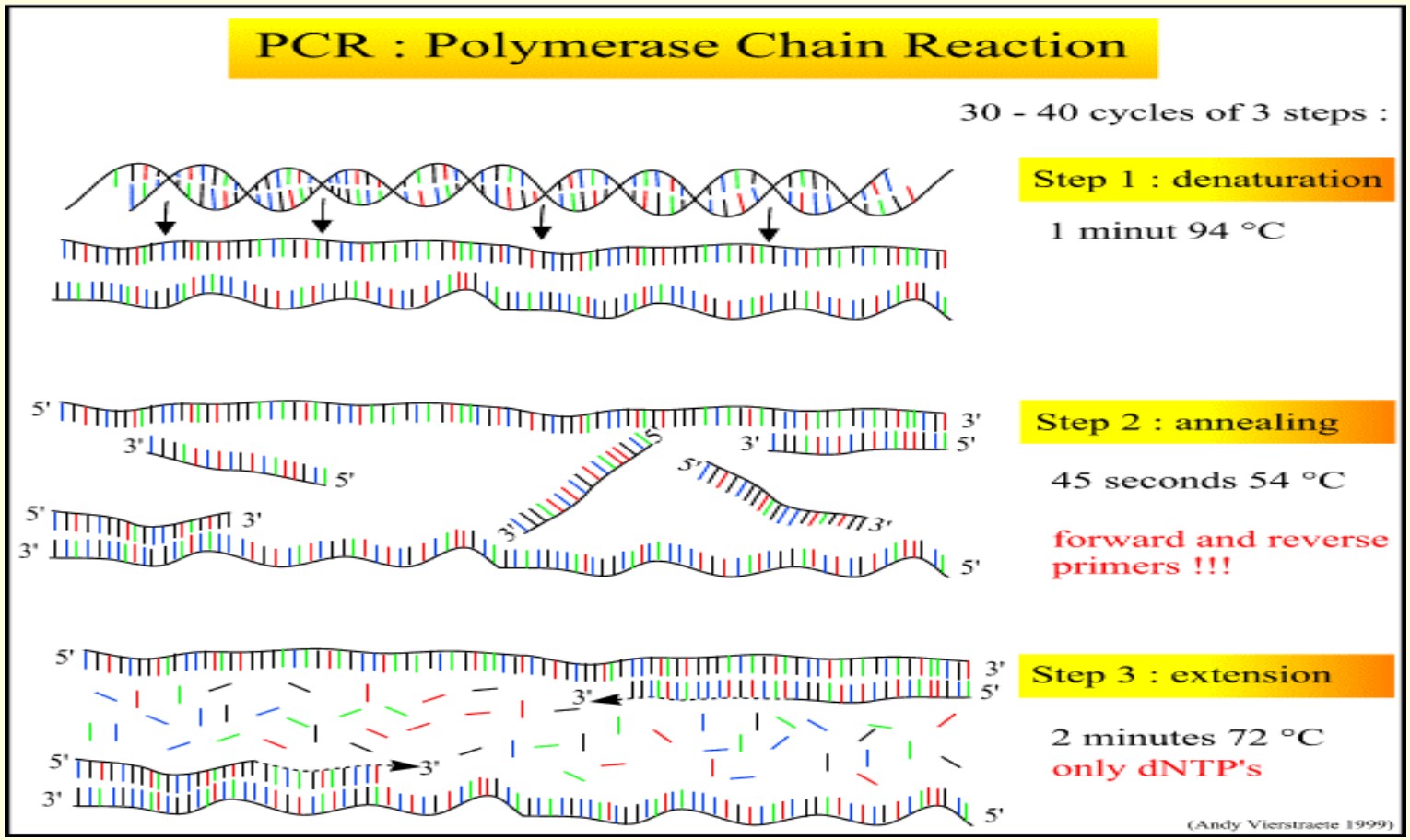
Degenerate Primer Design for Multiple Protein Alignment

Xintao Wei

11/26/2002

Introduction

- Polymerase chain reaction (PCR)



Primer

- Primer is a short, specific segments of DNA, providing specificity.

Template DNA

5’-ACATTTACGCGGATCCATAGGA……GGATCAGCATTACGATCCGGAAT- 3’

ACATTTACGCGGATCCAT

Forward primer

Reverse primer

TCGTAATGCTAGGCCTTA

3’-TGTAAATGCGCCTAGGTATCCT……..CCTAGTCGTAATGCTAGGCCTTA-5’

Degenerate PCR

- in most respect identical to ordinary PCR
- use mixed PCR primers ---- degenerate primers
- a very powerful tool to find "new" genes or gene families

Degenerate Primers

- **Degenerate primers** are multiple copies of an oligonucleotide, where in selected positions, the base is varied.

	Trp	Asp	Thr	Ala	Gly	Gln	Glu	
5’	**TGG**	**GAY**	**ACN**	**GCN**	**GGN**	**CAR**	**GAR**	**3’**
		GACACA	GATACA					
		GACACT	GATACT					
		GACACC	GATACC					
		GACACG	GATACG					

Where the Y=C or T, R=G or A, N=G, A, T or C.
Degeneracy = 2 * 4 * 4 * 4 * 2 * 2 = 256

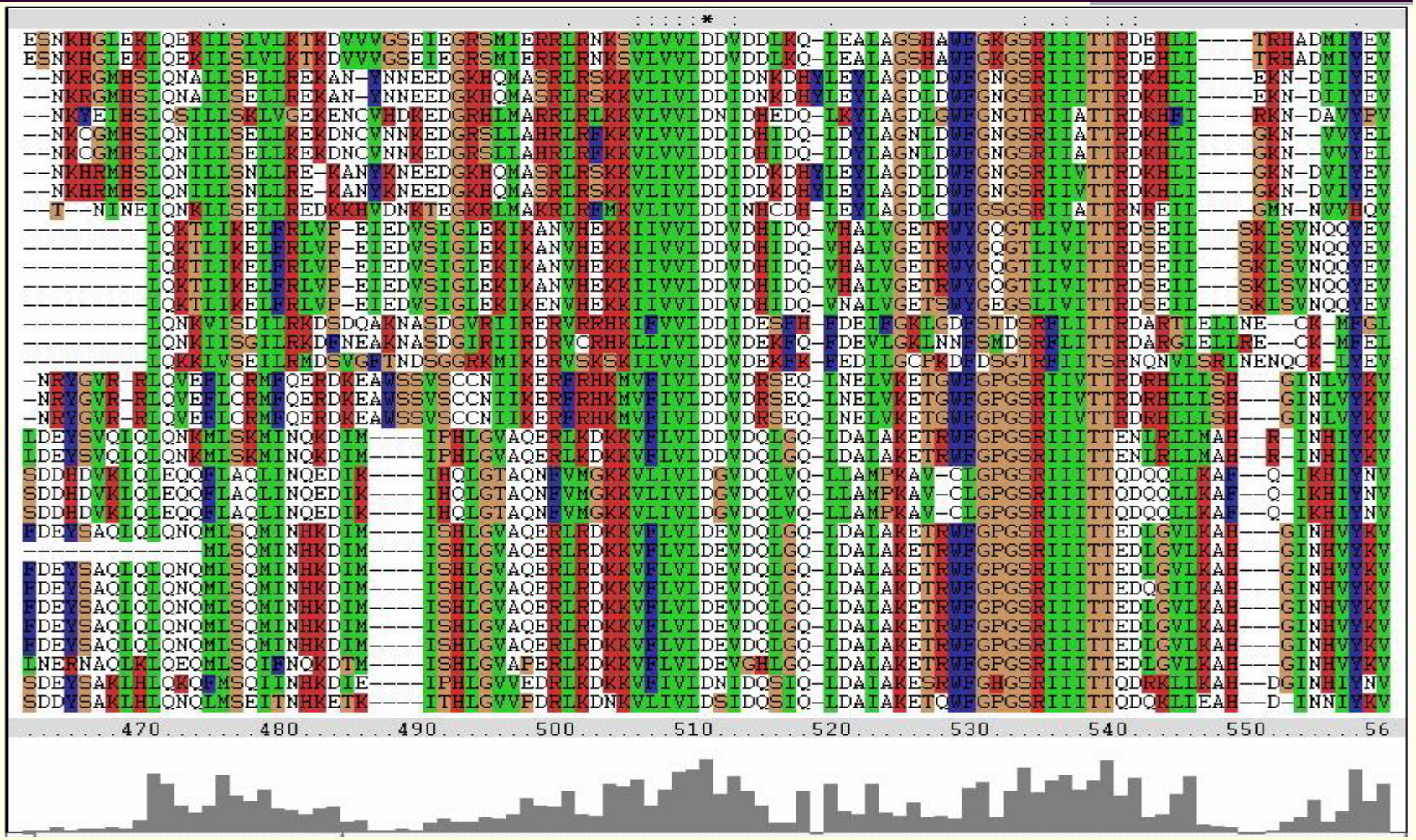
- The **degeneracy** of the primer is the number of unique sequence combinations it contains.

Example

- Toll/Interleukin-1 receptor homology (TIR): a characteristic domain of many plant resistance gene products

- An increasing number of TIR-encoding sequences are being identified through gene cloning, PCR amplification with degenerate primers, and genome sequencing projects.

TIRtop60 alignment



Applications

- Genefisher: Meyer,F., etc.
- CODEHOP: Ros,T.M., etc.

New Ideas

- Clustering
  - Group the sequences based on their sequence homology
  - In each group, find conserved regions

Program

- Program language: BioPerl
- Outline of the algorithm:
  - Build a “Block-score” matrix
  - Repeat until matrix is empty or biggest score < 2
    - Pick the biggest “Block-score”
    - Compute the conserved seq
    - If (can find primers)

      Put seqs into a group, and modify the matrix.
    - Else

      Output the seqs as separate groups

      Modify the matrix
  - Output all seqs as groups

Result

- 6 groups
- Example: group #1
  - The number of sequences is 11

Conserved region	position	Degenerate primers	degeneracy
GPGSRIIITT	530	GGTCCTGGRWSTCGKATTATY ATCACAACM	64
EA[FL]QIFC	566	GAGGCYYTTCAAATCTTYTGY	16
Q[FL]GRE[TI][SV]	741	CARYTBGGWAGAGAAAYTKYY	384