Exciting the Reluctant Bioinformatician

Limsoon Wong

NUS National University of Singapore

Plan

- **NUS Bioinformatics Programme**

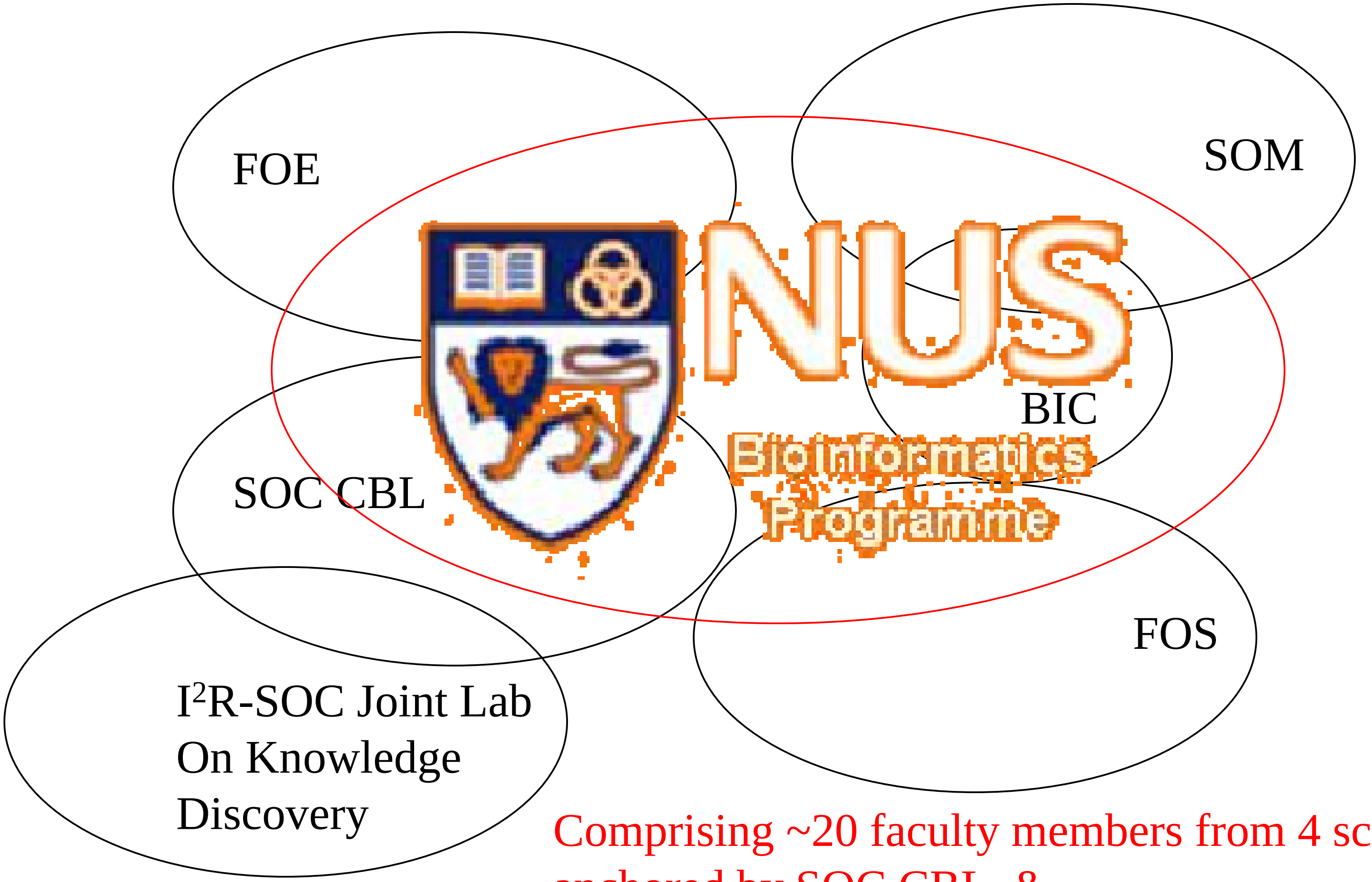
- **Research**
  - Themes
  - Collaborations
  - Some Basic Bioinformatics Results in 2006

- **Education**
  - Core courses
  - Key principles emphasized in first course

NUS Bioinformatics Programme

NUS
National University
of Singapore

Structure



FOE

SOM

BIC

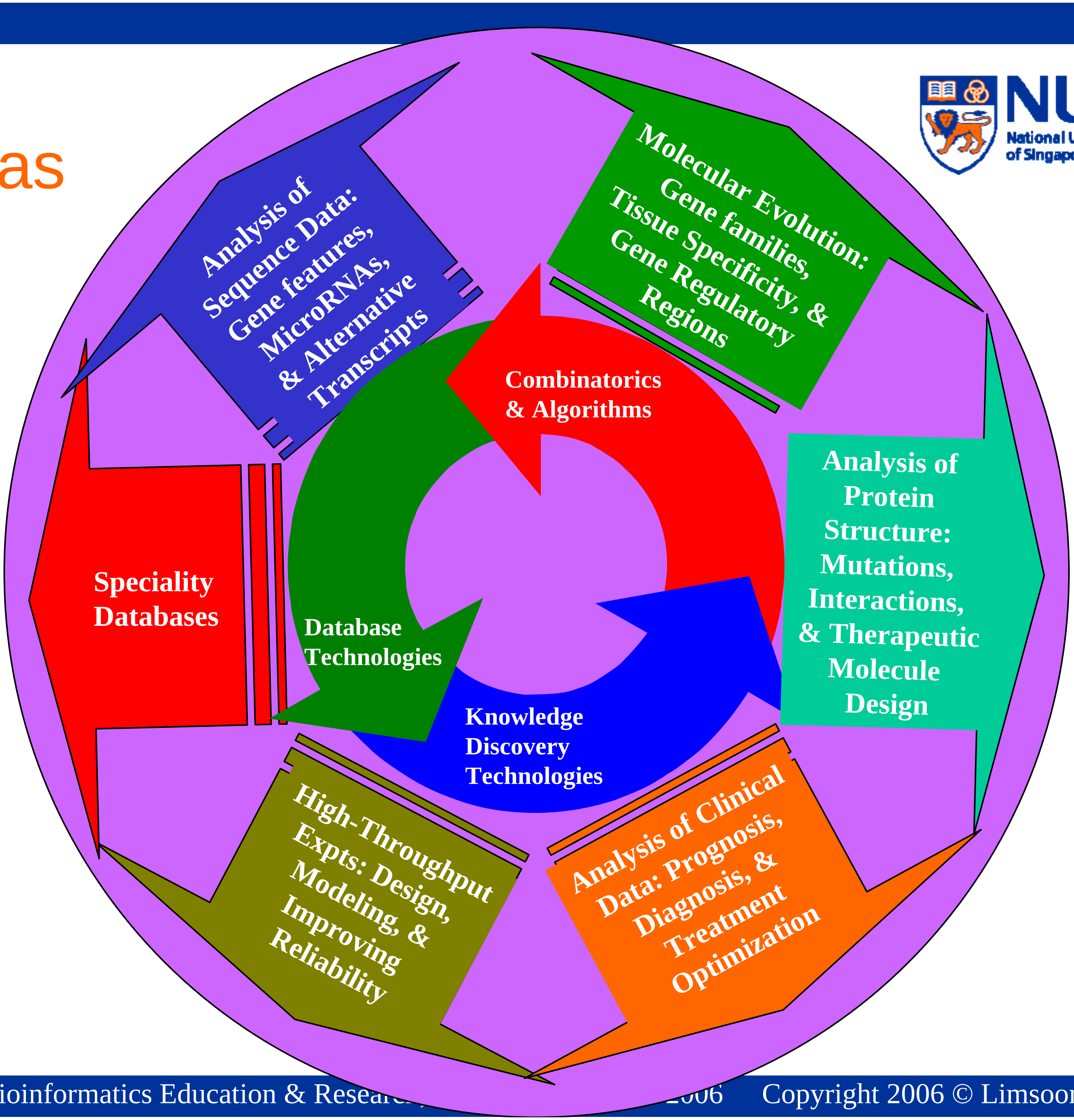
SOC CBL

FOS

I^{2}R-SOC Joint Lab On Knowledge Discovery

Comprising ~20 faculty members from 4 schools, anchored by SOC CBL, & coordinated by Limsoon Wong

Areas



Analysis of Sequence Data: Gene features, MicroRNAs, & Alternative Transcripts

Molecular Evolution: Gene families, Tissue Specificity, & Gene Regulatory Regions

Speciality Databases

Combinatorics & Algorithms

Database Technologies

Knowledge Discovery Technologies

Analysis of Protein Structure: Mutations, Interactions, & Therapeutic Molecule Design

High-Throughput Expts: Design, Modeling, & Improving Reliability

Analysis of Clinical Data: Prognosis, Diagnosis, & Treatment Optimization

People

- **School of Computing**

Ken Sung

Anthony Tung

Mong Li Lee

Wynne Hsu

HweeTou Ng

Kian Lee Tan

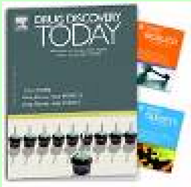
Hon Wai Leong

Limsoon Wong

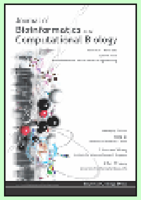
- **School of Medicine**
  - Robert HEWITT, Coral LAI, SK SETHI, Tin Wee TAN, Bor-Luen TANG, Allen YEOH
- **Faculty of Engineering**
  - Dong Yup LEE, Hai LIN
- **Faculty of Science**
  - Jinhua HAN, Yong KONG, Susan MOORE, Martti TAMMI, Louxin ZHANG
- **Staff**
  - Mark DE SILVA, Kuan Siong LIM

Professional Activities in 2005/6

- **Journal edited:**

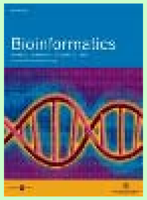


DDT



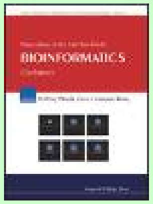
JBCB

IJBRA

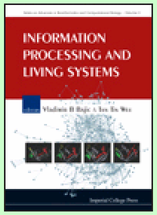


Bioinformatics

- **Books/Proceedings edited:**



APBC'05



Information Processing & Living Systems

APBC'06

PRIB'06

- **Involved in 20+ bioinformatics conf prog & org committees**
  - APBC05, APBC06, CSB05, CSB06, ECCB05, GIW05, GIW06, ISMB05, ISMB06, PSB06, …
- **Published 100+ papers**
  - Bioinformatics, JCB, BMC, JBCB, Bioinformatics, Nature Methods, NAR, Mol Biol Cell, Hum Mol Genet, Metab Eng, …
- **20+ keynotes & invited talks in conferences**

Conferences Hosted in 2006

- **5th Korea-Singapore Workshop on Bioinformatics and NLP**
  - Feb 2006 @ NUS SOC
- **IMS Workshop on BioAlgorithmics**
  - July 2006 @ NUS IMS
- **3rd RECOMB Satellite Workshop on Regulatory Genomics**
  - July 2006 @ NUS SOC
- **Forthcoming:**
  - LBM2007, AASBi2007, GIW2007, RECOMB2008

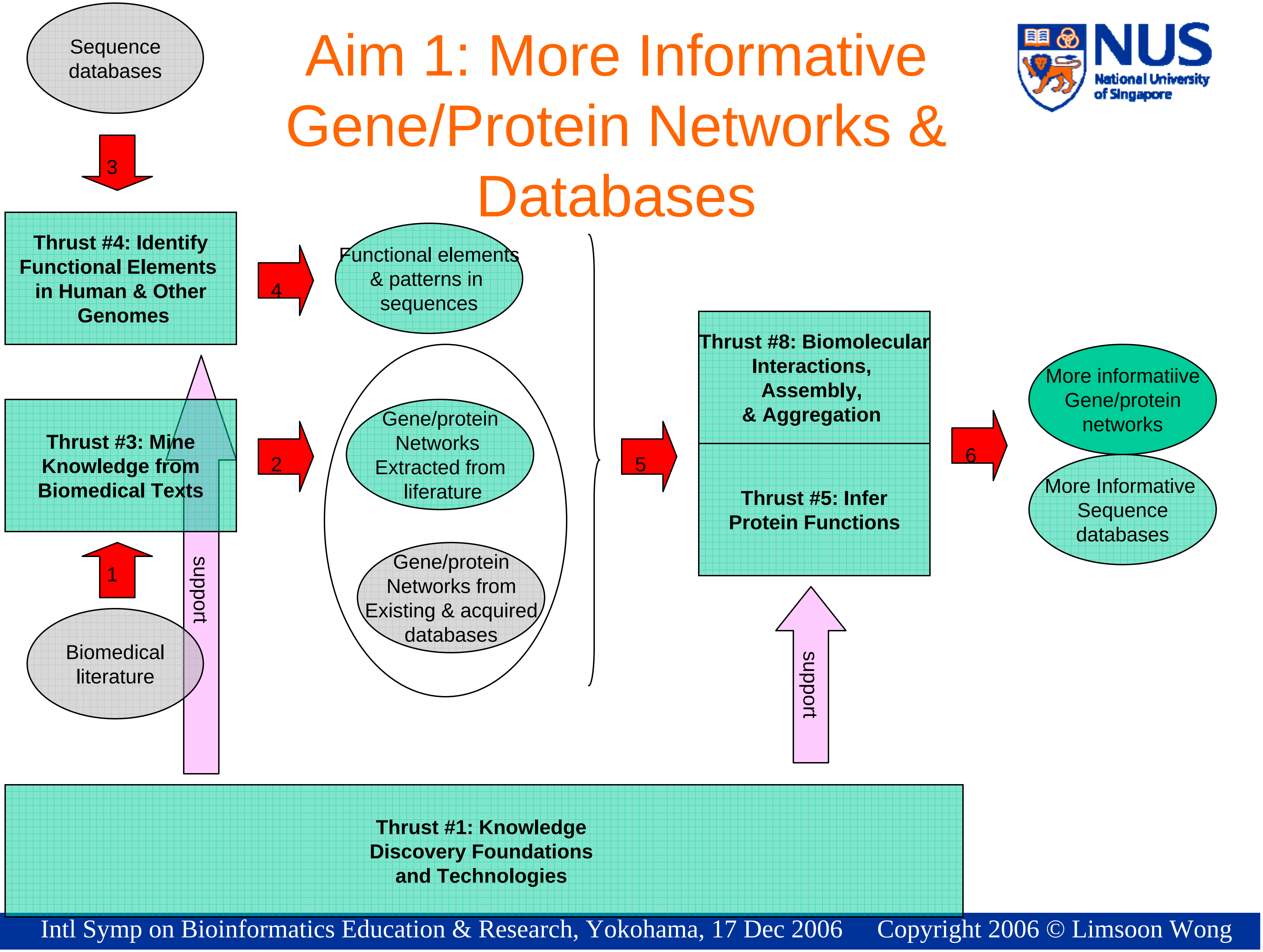
Honours

- **Ken Sung**
  - 2006 Singapore National Science Award: Paired End diTag sequencing technology
  - 2003 Japan Forum on IT Award: Space-efficient algo for full-text indices

- **Limsoon Wong**
  - 2006 Singapore Youth Award Medal of Commendation: Sustained contributions to science & technology
  - 2003 Far Eastern Economic Review Asian Innovation Gold Award: A simple test for childhood leukaemia

Research

Aim 1: More Informative Gene/Protein Networks & Databases



Sequence databases

3

Thrust #4: Identify Functional Elements in Human & Other Genomes

4

Functional elements & patterns in sequences

Biomedical literature

1

Thrust #3: Mine Knowledge from Biomedical Texts

support

2

Gene/protein Networks Extracted from liferature

Gene/protein Networks from Existing & acquired databases

5

Thrust #8: Biomolecular Interactions, Assembly, & Aggregation

Thrust #5: Infer Protein Functions

support

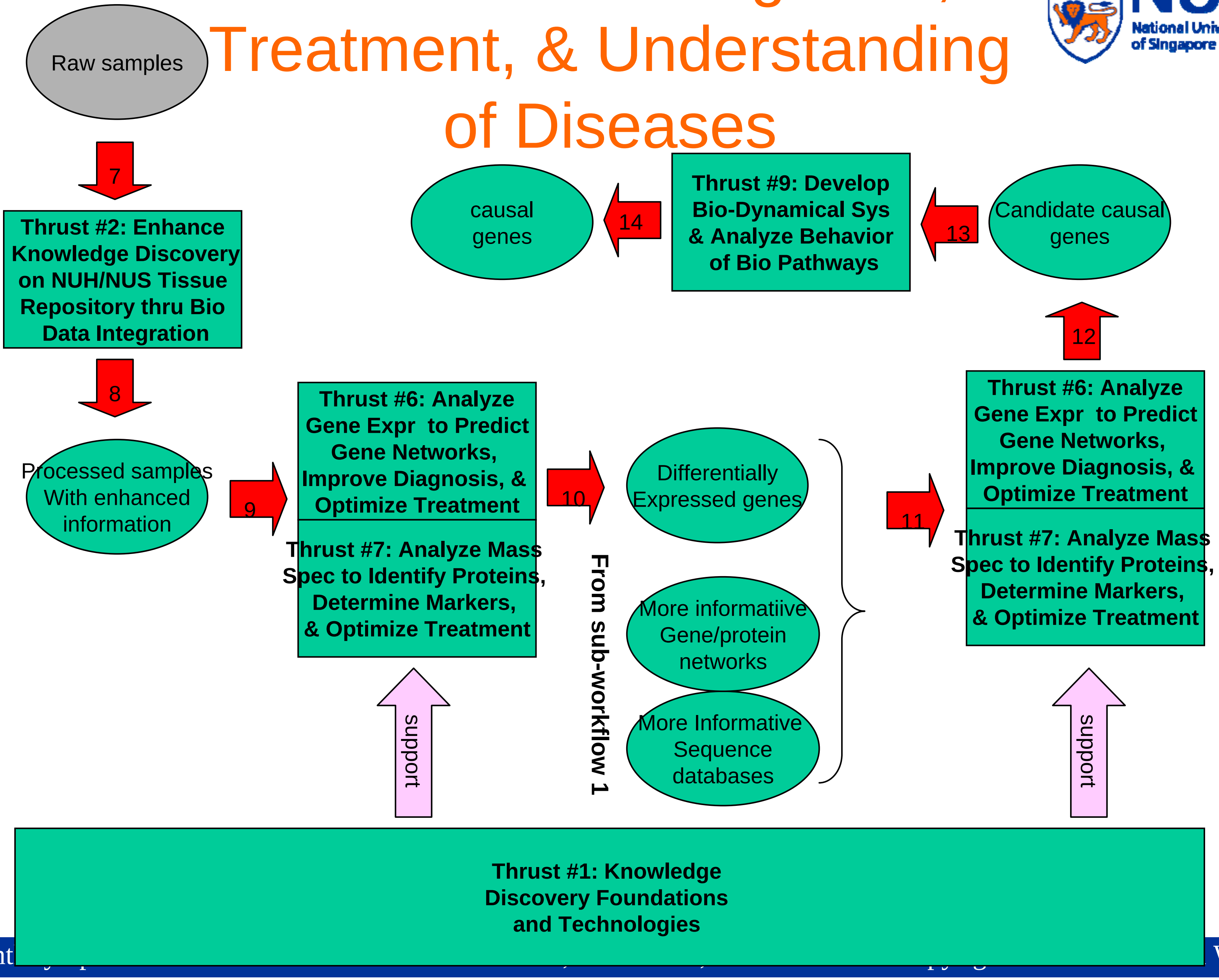
6

More informatiive Gene/protein networks

More Informative Sequence databases

Thrust #1: Knowledge Discovery Foundations and Technologies

Aim 2: Better Diagnosis, Treatment, & Understanding of Diseases



Raw samples

7

Thrust #2: Enhance Knowledge Discovery on NUH/NUS Tissue Repository thru Bio Data Integration

8

Processed samples With enhanced information

9

Thrust #6: Analyze Gene Expr to Predict Gene Networks, Improve Diagnosis, & Optimize Treatment

Thrust #7: Analyze Mass Spec to Identify Proteins, Determine Markers, & Optimize Treatment

support

10

From sub-workflow 1

Differentially Expressed genes

More informatiive Gene/protein networks

More Informative Sequence databases

11

Thrust #6: Analyze Gene Expr to Predict Gene Networks, Improve Diagnosis, & Optimize Treatment

Thrust #7: Analyze Mass Spec to Identify Proteins, Determine Markers, & Optimize Treatment

support

12

Candidate causal genes

13

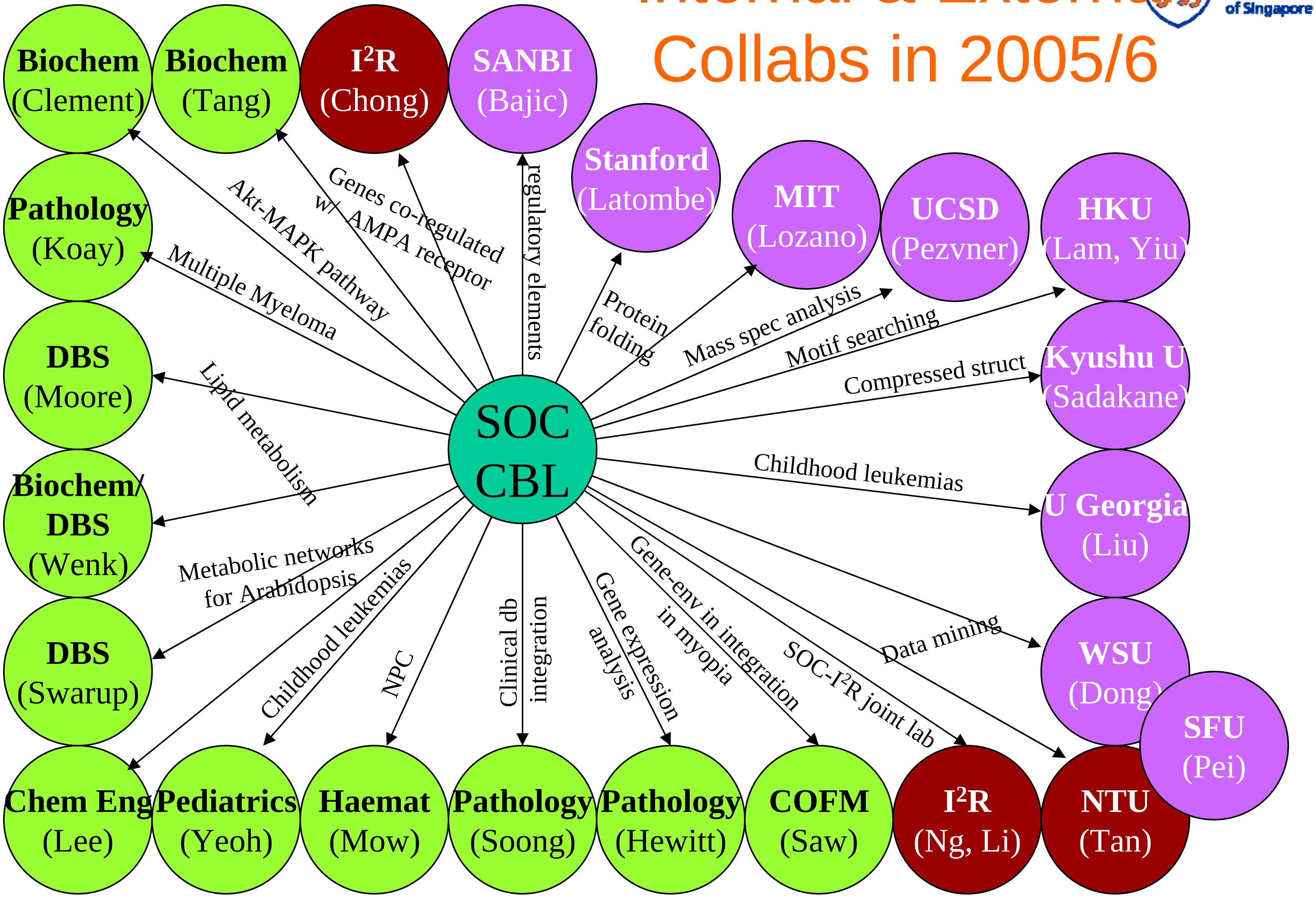
Thrust #9: Develop Bio-Dynamical Sys & Analyze Behavior of Bio Pathways

14

causal genes

Thrust #1: Knowledge Discovery Foundations and Technologies

Internal & External Collabs in 2005/6



SOC CBL

Biochem (Clement)

Akt-MAPK pathway

Biochem (Tang)

Genes co-regulated w/ AMPA receptor

I^{2}R (Chong)

SANBI (Bajic)

regulatory elements

Stanford (Latombe)

Protein folding

MIT (Lozano)

Mass spec analysis

UCSD (Pezvner)

Motif searching

HKU (Lam, Yiu)

Compressed struct

Kyushu U (Sadakane)

Childhood leukemias

U Georgia (Liu)

Data mining

WSU (Dong)

SFU (Pei)

NTU (Tan)

SOC-I^{2}R joint lab

I^{2}R (Ng, Li)

Gene-env in integration in myopia

COFM (Saw)

Gene expression analysis

Pathology (Hewitt)

Clinical db integration

Pathology (Soong)

NPC

Haemat (Mow)

Childhood leukemias

Pediatrics (Yeoh)

Chem Eng (Lee)

Metabolic networks for Arabidopsis

DBS (Swarup)

Biochem/ DBS (Wenk)

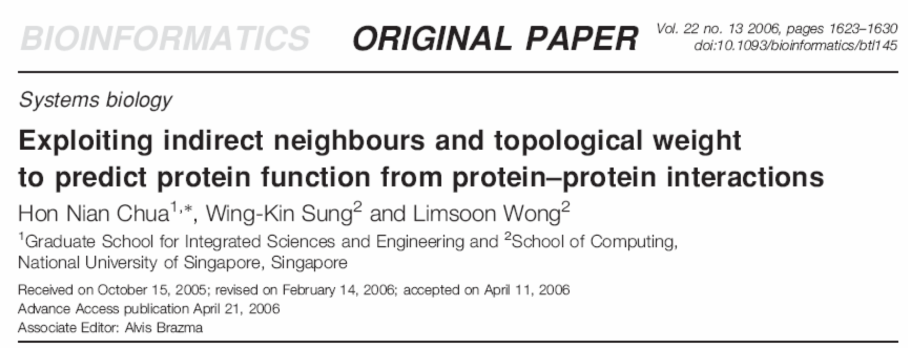
Lipid metabolism

DBS (Moore)

Multiple Myeloma

Pathology (Koay)

Protein Function Prediction:
A Central Problem in Computational Biology



BIOINFORMATICS ORIGINAL PAPER Vol. 22 no. 13 2006, pages 1623–1630
doi:10.1093/bioinformatics/btl145

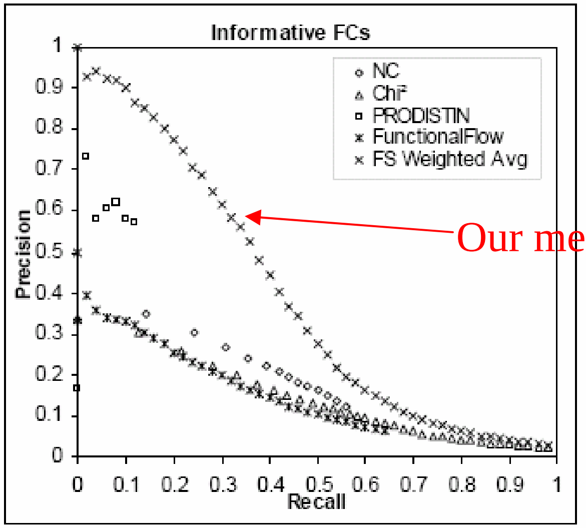
Systems biology

Exploiting indirect neighbours and topological weight to predict protein function from protein–protein interactions

Hon Nian Chua[1,*], Wing-Kin Sung[2] and Limsoon Wong[2]

[1]Graduate School for Integrated Sciences and Engineering and [2]School of Computing, National University of Singapore, Singapore

Received on October 15, 2005; revised on February 14, 2006; accepted on April 11, 2006
Advance Access publication April 21, 2006
Associate Editor: Alvis Brazma



Our method

This project is supported in part by a A*STAR AGS scholarship,
and the I^{2}R-SOC Joint Lab on Knowledge Discovery fromClinical Data

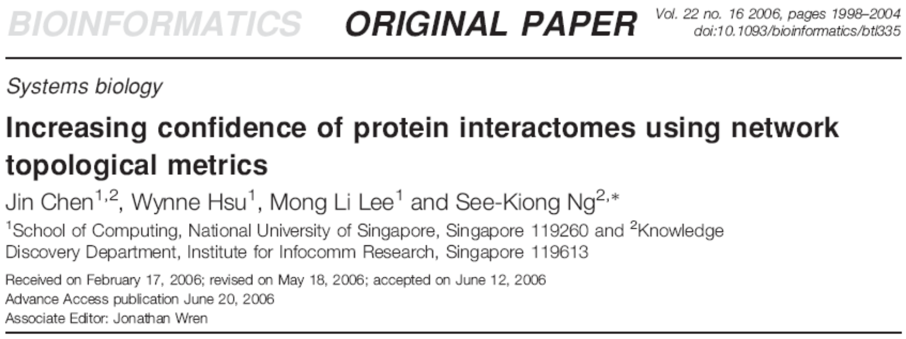
- How significant is functional association between level-2 neighbors?
- How can they be exploited for protein function prediction?
- How to integrate protein interaction info with other info to improve protein function prediction?

⇒ **Robust and powerful system to predict protein functions, even w/o sequence homology**

Protein Interactions Reliability: A Bottleneck in Proteomic Research

- **Protein-protein interaction expts have ~50% errors**
- **True interactions seem to exhibit certain topologies and motifs that can be modeled**
- **Develop computational methods to detect false positives**
- **Develop computational methods to detect false negatives**

⇒ **Robust and powerful system to identify protein-protein interactions in noisy expts**



BIOINFORMATICS ORIGINAL PAPER Vol. 22 no. 16 2006, pages 1998–2004
doi:10.1093/bioinformatics/btl335

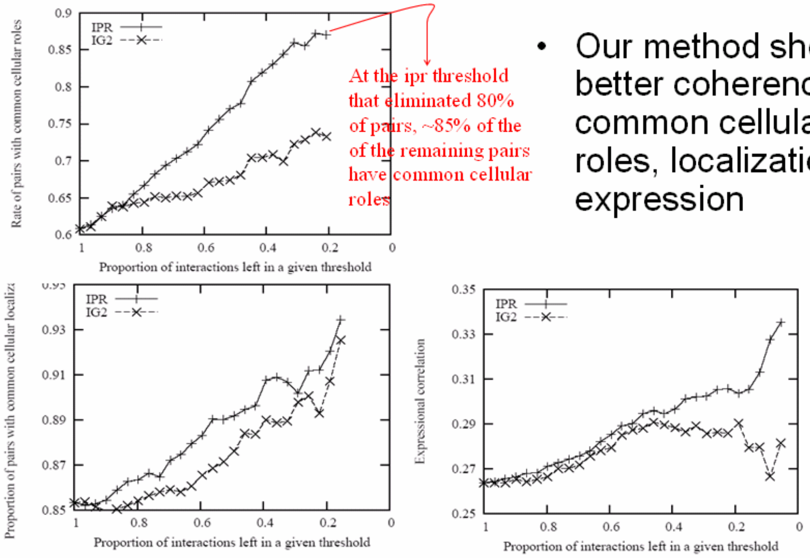
Systems biology

Increasing confidence of protein interactomes using network topological metrics

Jin Chen[1,2], Wynne Hsu[1], Mong Li Lee[1] and See-Kiong Ng[2,*]

[1]School of Computing, National University of Singapore, Singapore 119260 and [2]Knowledge Discovery Department, Institute for Infocomm Research, Singapore 119613

Received on February 17, 2006; revised on May 18, 2006; accepted on June 12, 2006
Advance Access publication June 20, 2006
Associate Editor: Jonathan Wren

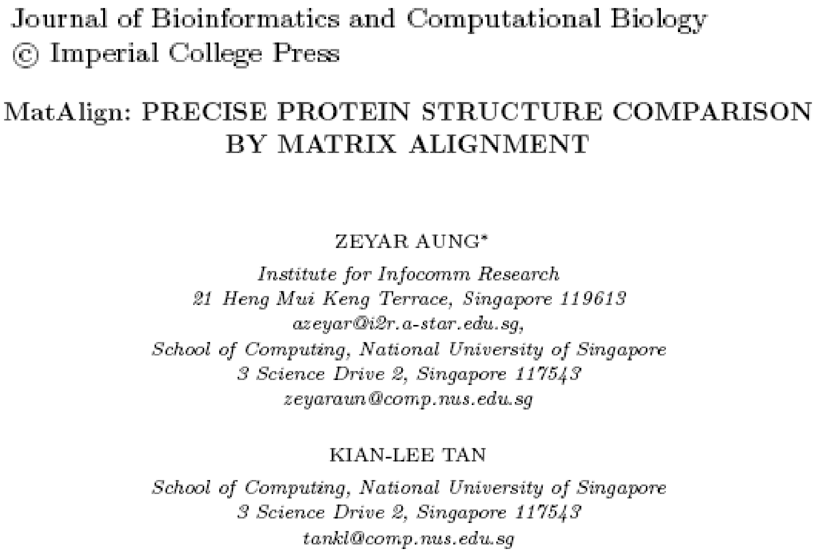


At the ipr threshold that eliminated 80% of pairs, ~85% of the of the remaining pairs have common cellular roles

- Our method shows better coherence wrt common cellular roles, localization, & expression

This project is supported in part by the I^{2}R-SOC Joint Lab on Knowledge Discovery from Clinical Data

Precise Structure Comparison: A Key Problem in Structural Bioinformatics

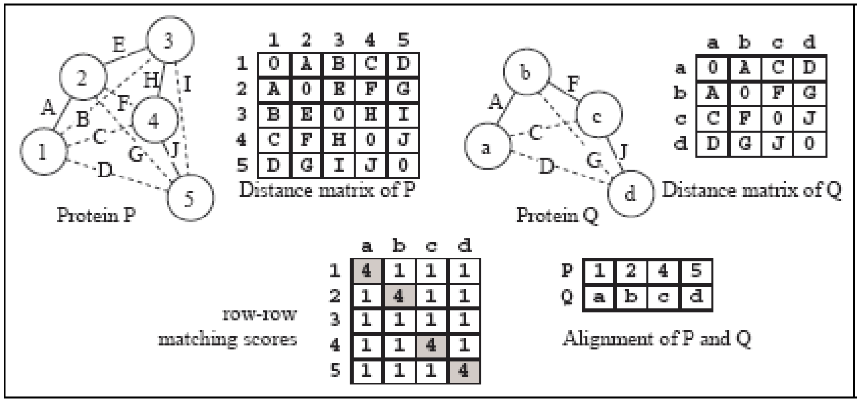


Journal of Bioinformatics and Computational Biology
© Imperial College Press

MatAlign: PRECISE PROTEIN STRUCTURE COMPARISON BY MATRIX ALIGNMENT

ZEYAR AUNG*
Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
azeyar@i2r.a-star.edu.sg,
School of Computing, National University of Singapore
3 Science Drive 2, Singapore 117543
zeyaraun@comp.nus.edu.sg

KIAN-LEE TAN
School of Computing, National University of Singapore
3 Science Drive 2, Singapore 117543
tankl@comp.nus.edu.sg



	1	2	3	4	5
1	0	A	B	C	D
2	A	0	E	F	G
3	B	E	0	H	I
4	C	F	H	0	J
5	D	G	I	J	0

Distance matrix of P

	a	b	c	d
a	0	A	C	D
b	A	0	F	G
c	C	F	0	J
d	D	G	J	0

Distance matrix of Q

	a	b	c	d
1	4	1	1	1
2	1	4	1	1
3	1	1	1	1
4	1	1	4	1
5	1	1	1	4

P	1	2	4	5
Q	a	b	c	d

Alignment of P and Q

- **MatAlign**
  - Detailed struct alignment thru alignment of 2D dist matrix & iterative refinements
  - Provide better alignment scores than DALI & CE in majority of cases
  - 4 times faster than DALI, and has about the same speed as CE

⇒ **Significantly speed up searching of protein sequences and structures w/o sacrificing accuracy**

Education

Main Courses Developed

- **CS2220 Introduction to Computational Biology**
  - Understand bioinformatics problems; interpretational skills
- **CS3225 Combinatorial Methods in Bioinformatics**
- **CS4220 Knowledge Discovery Methods in Bioinformatics**
  - Clustering; classification; association rules; SVM; HMM; Mining of seq, trees, & graphs
- **CS5238 Advanced Combinatorial Methods in Bioinformatics**
  - Seq alignment, whole-genome alignment, suffix tree, seq indexing, motif finding, RNA sec struct prediction, phylogeny reconstruction
- **CS6280 Computational Systems Biology**
  - Dynamics of biochemical and signaling networks; modeling, simulating, & analyzing them
- **Etc …**

Things Taught in CS2220: Our “Intro to Bioinformatics” Course

- **Tastes of Bioinformatics Problems**
  - Multi-step nature
  - Noisy biased data
- **Core principles**
  - Guilt by Association
  - Emerging Patterns
  - Identifying and Exploiting “Invariants”
- **Techniques**
  - Knowledge discovery methodology
  - Very basic knowledge discovery methods
  - Very basic combinatorial methods

A running example based on protein function prediction …

SPSTNRKYPPLPVDKLEEEINRRMADDNKLFREEFNALPACPIQATCEAASKEENKEKNR
YVNILPYDHSRVHLTPVEGVPDSDYINASFINGYQEKNKFIAAQGPKEETVNDFWRMIWE
QNTATIVMVTNLKERKECKCAQYWPDQGCWTYGNVRVSVEDVTVLVDYTVRKFCIQQVGD
VTNRKPQRLITQFHFTSWPDFGVPFTPIGMLKFLKKVKACNPQYAGAIVVHCSAGVGRTG
TFVVIDAMLDMMHSERKVDVYGFVSRIRAQRCQMVQTDMQYVFIYQALLEHYLYGDTELE
VT

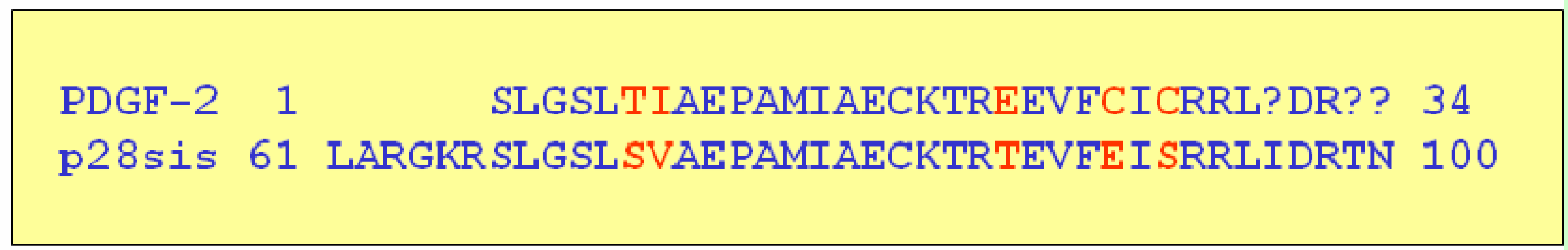
- **How do we attempt to assign a function to a new protein sequence?**

Guilt by Association

(General Idea & Many Manifestations)

- Compare the target sequence T with sequences $S_1, \ldots, S_n$ of known function in a db
- Determine which ones amongst $S_1, \ldots, S_n$ are the likely homologs of T
- Then assign to T the same function as these homologs
- Finally, confirm with suitable wet experiments

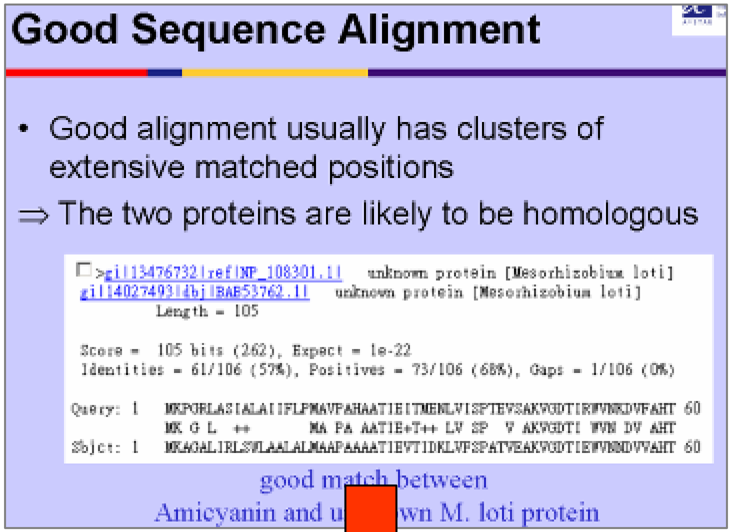
Guilt by Association of Sequence Similarity



```
PDGF-2  1       SLGSLTIAEPAMIAECKTREEVFCICRRL?DR?? 34
p28sis 61 LARGKRSLGSLSVAEPAMIAECKTRTEVFEISRRLIDRTN 100
```

Guilt by Association of Seq Similarity

Compare *T* with seqs of known function in a db



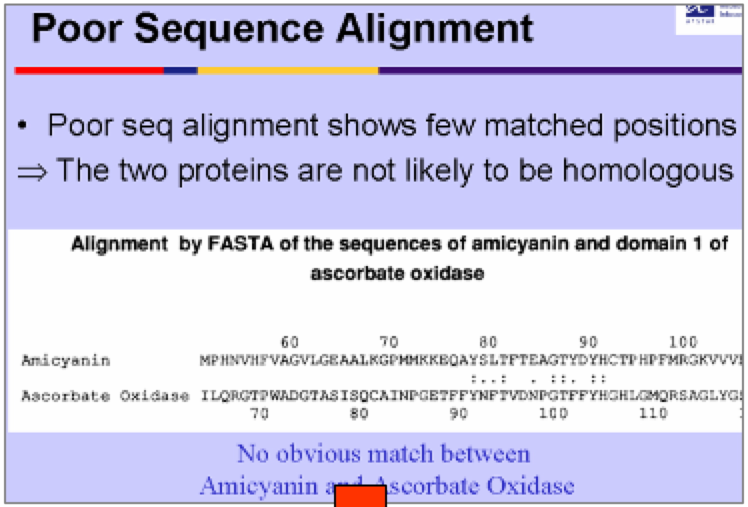
Good Sequence Alignment

- Good alignment usually has clusters of extensive matched positions

⇒ The two proteins are likely to be homologous

```
□>gi|13476732|ref|NP_108301.1|    unknown protein [Mesorhizobium loti]
 gi|14027493|dbj|BAB53762.1|   unknown protein [Mesorhizobium loti]
          Length = 105

 Score =  105 bits (262), Expect = 1e-22
 Identities = 61/106 (57%), Positives = 73/106 (68%), Gaps = 1/106 (0%)

Query: 1    MKPGRLASIALAIIFLPMAVPAHAATIEITMENLVISPTEVSAKVGDTIRWVNKDVFAHT  60
             MK G L  ++         MA PA AATIE+T++ LV SP   V AKVGDTI WVN DV AHT
Sbjct: 1    MKAGALIRLSWLAALALMAAPAAAATIEVTIDKLVFSPATVEAKVGDTIEWVNNDVVAHT  60
```

good m tch between
Amicyanin and u wn M. loti protein



Poor Sequence Alignment

- Poor seq alignment shows few matched positions

⇒ The two proteins are not likely to be homologous

Alignment by FASTA of the sequences of amicyanin and domain 1 of ascorbate oxidase

```
                          60        70        80        90        100
Amicyanin          MPHNVHFVAGVLGEAALKGPMMKKEQAYSLTFTEAGTYDYHCTPHPFMRGKVVV
                                                :..:   . ::. ::
Ascorbate Oxidase  ILQRGTPWADGTASISQCAINPGETFFYNFTVDNPGTFFYHGHLGMQRSAGLYG
                         70        80        90        100       110
```

No obvious match between
Amicyanin a   scorbate Oxidase

Discard this function as a candidate

Assign to *T* same function as homologs

Confirm with suitable wet experiments

What if there is no useful homolog? Guilt by other types of association!

- **Similarity of dissimilarities (e.g., SVM-PAIRWISE)**
- **Similarity of phylogenetic profiles**
- **Similarity of subcellular co-localization & other physico-chemico properties (e.g., PROTFUN)**
- **Similarity of gene expression profiles**
- **Similarity of protein-protein interaction partners**
- **…**

Guilt by Association of Similarity of Dissimilarities

Image credit: www.comstock.com

Similarity of Dissimilarities

	orange$_1$	banana$_1$	…

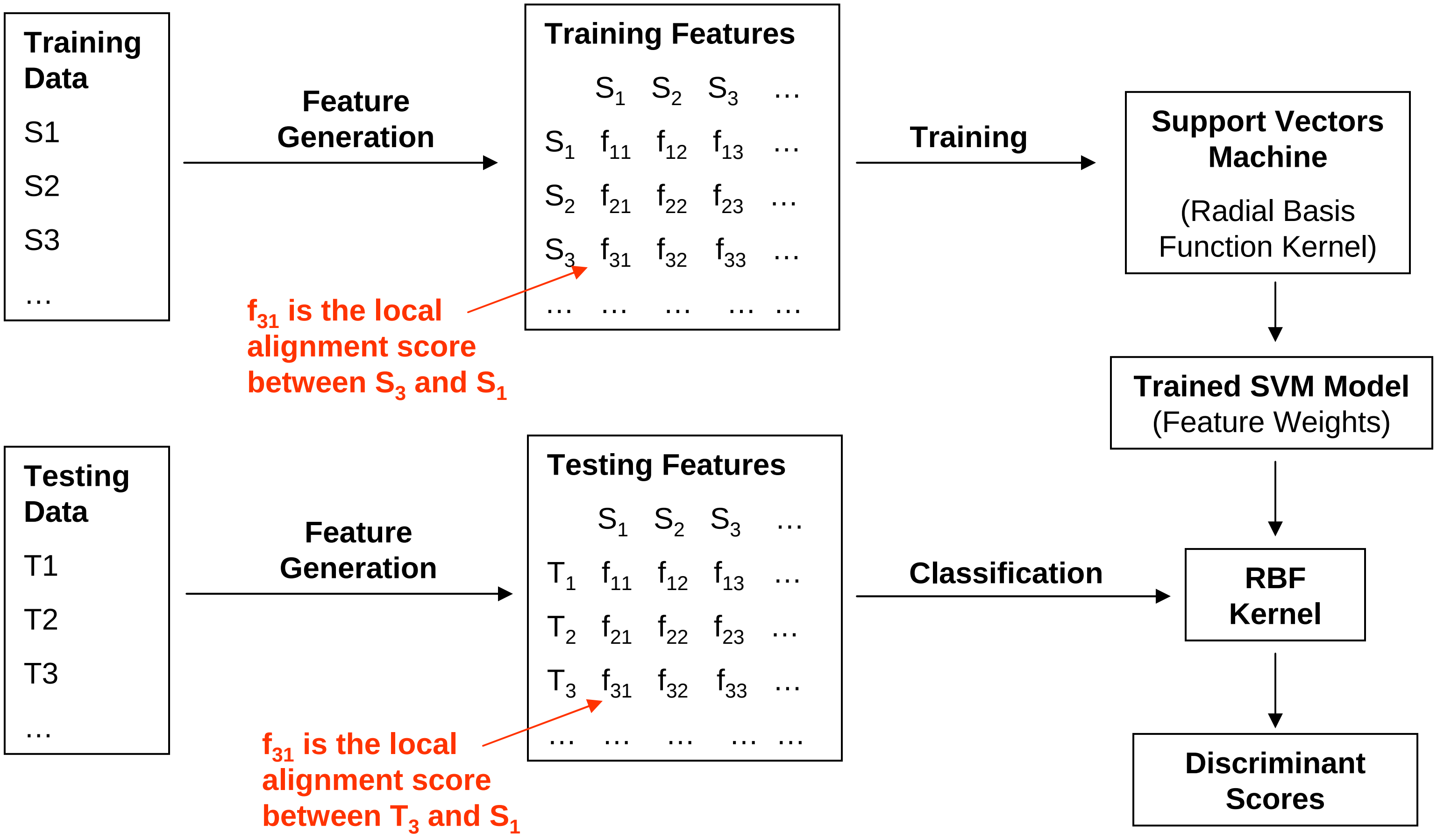
apple$_1$	Color = red vs orange
	Skin = smooth vs rough
	Size = small vs small
	Shape = round vs round
		Color = red vs yellow
		Skin = smooth vs smooth
		Size = small vs small
		Shape = round vs oblong
			…

apple$_2$	Color = red vs orange
	Skin = smooth vs rough
	Size = small vs small
	Shape = round vs round
		Color = red vs yellow
		Skin = smooth vs smooth
		Size = small vs small
		Shape = round vs oblong
			…

orange$_2$	Color = orange vs orange
	Skin = rough vs rough
	Size = small vs small
	Shape = round vs round
		Color = orange vs yellow
		Skin = rough vs smooth
		Size = small vs small
		Shape = round vs oblong
			..

…	…	…	…

SVM-Pairwise Framework



Training Data

S1

S2

S3

…

Feature Generation

Training Features

	S_1	S_2	S_3	…
S_1	f_{11}	f_{12}	f_{13}	…
S_2	f_{21}	f_{22}	f_{23}	…
S_3	f_{31}	f_{32}	f_{33}	…
…	…	…	…	…

f_{31} is the local alignment score between S_3 and S_1

Training

Support Vectors Machine

(Radial Basis Function Kernel)

Trained SVM Model
(Feature Weights)

Testing Data

T1

T2

T3

…

Feature Generation

Testing Features

	S_1	S_2	S_3	…
T_1	f_{11}	f_{12}	f_{13}	…
T_2	f_{21}	f_{22}	f_{23}	…
T_3	f_{31}	f_{32}	f_{33}	…
…	…	…	…	…

f_{31} is the local alignment score between T_3 and S_1

Classification

RBF Kernel

Discriminant Scores

Image credit: Kenny Chua

Guilt by Association of Genome Phylogenetic Profiles

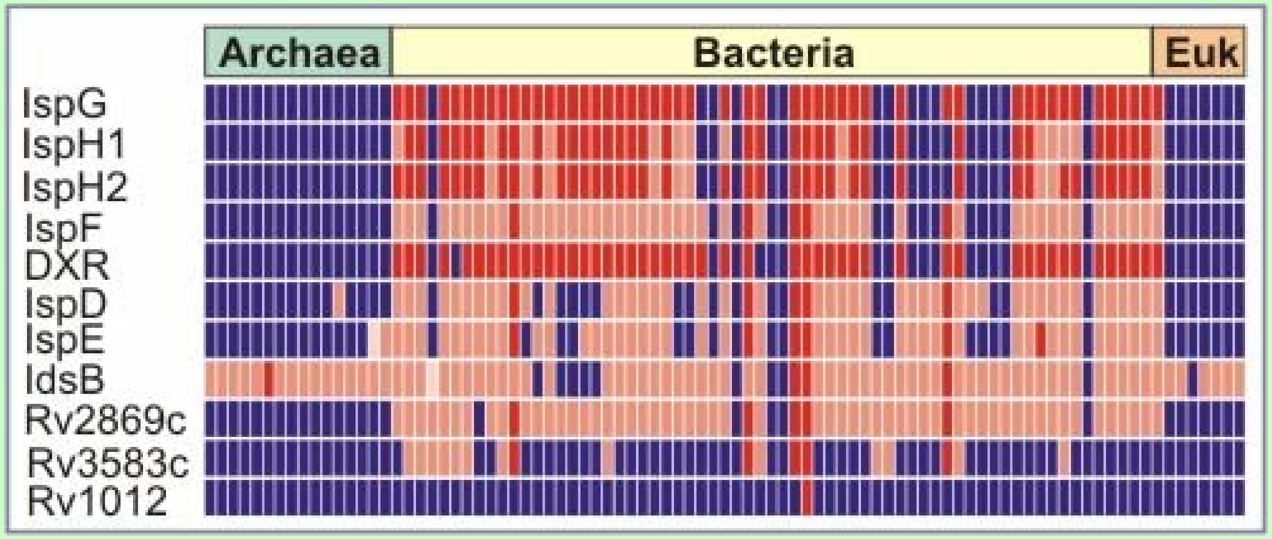


Image credit: Ed Marcotte, http://apropos.icmb.utexas.edu/plex/tour/isoprenoid.jpg

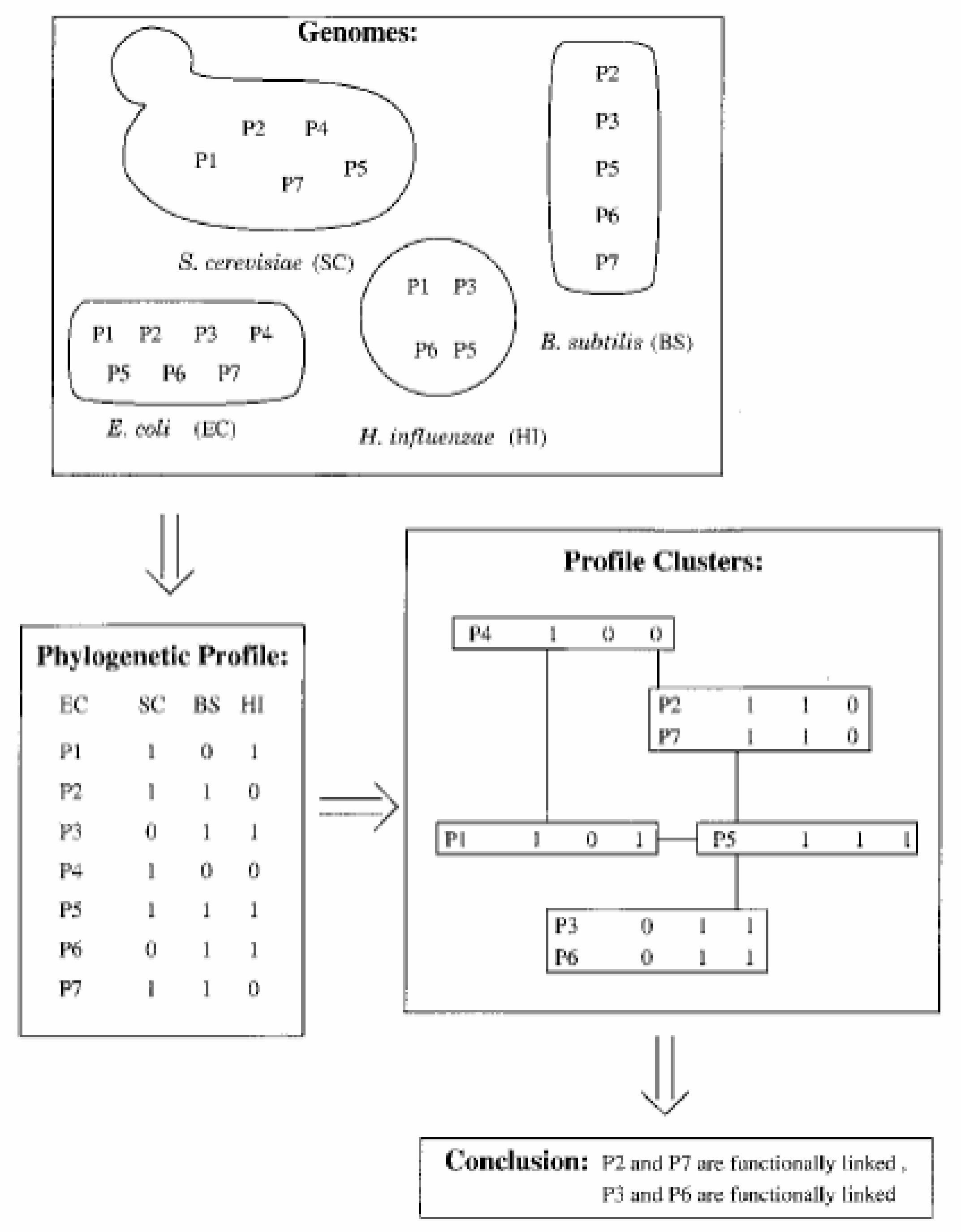
Phylogenetic Profiling

Pellegrini et al., *PNAS*, 96:4285--4288, 1999

- **Gene (and hence proteins) with identical patterns of occurrence across phyla tend to function together**

⇒ **Even if no homolog with known function is available, it is still possible to infer function of a protein**

Phylogenetic Profiling: How It Works



Phylogenetic Profile:

EC	SC	BS	HI
P1	1	0	1
P2	1	1	0
P3	0	1	1
P4	1	0	0
P5	1	1	1
P6	0	1	1
P7	1	1	0

Conclusion: P2 and P7 are functionally linked ,
P3 and P6 are functionally linked

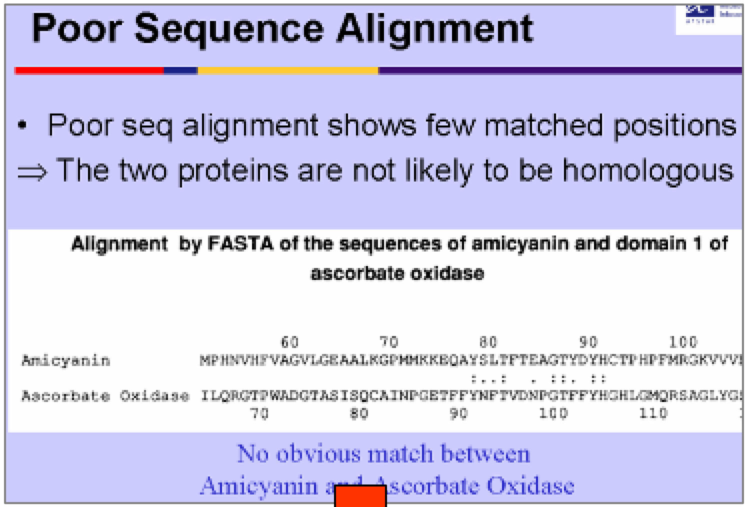
Twists in the Tale of Guilt by Association of Seq Similarity

(Noisy & Biased Data)

Image credit: Shanti Christensen, http://static.flickr.com/46/148437681_7f2dfa977e_m.jpg

Guilt by Association of Seq Similarity

Compare *T* with seqs of known function in a db



Poor Sequence Alignment

- Poor seq alignment shows few matched positions

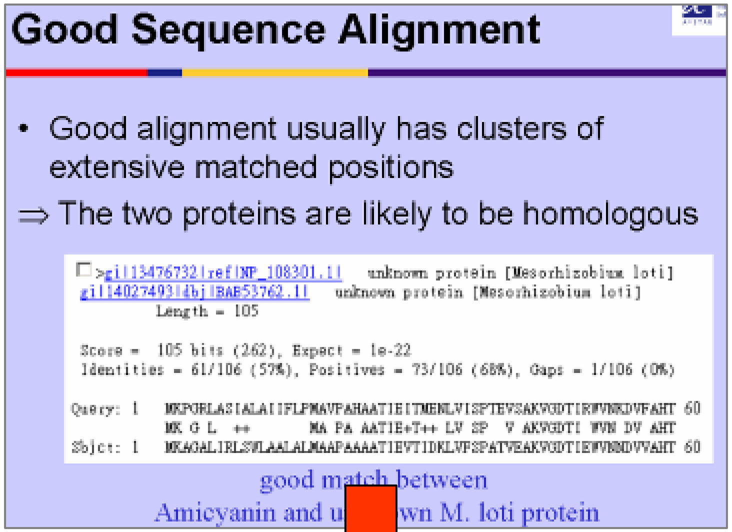
⇒ The two proteins are not likely to be homologous

Alignment by FASTA of the sequences of amicyanin and domain 1 of ascorbate oxidase

```
                        60        70        80        90        100
Amicyanin          MPHNVHFVAGVLGEAALKGPMMKKEQAYSLTFTEAGTYDYHCTPHPFMRGKVVV
                                                 :..:   . ::. ::
Ascorbate Oxidase  ILQRGTPWADGTASISQCAINPGETFFYNFTVDNPGTFFYHGHLGMQRSAGLYG
                      70        80        90        100       110
```

No obvious match between Amicyanin and Ascorbate Oxidase

Discard this function as a candidate



Good Sequence Alignment

- Good alignment usually has clusters of extensive matched positions

⇒ The two proteins are likely to be homologous

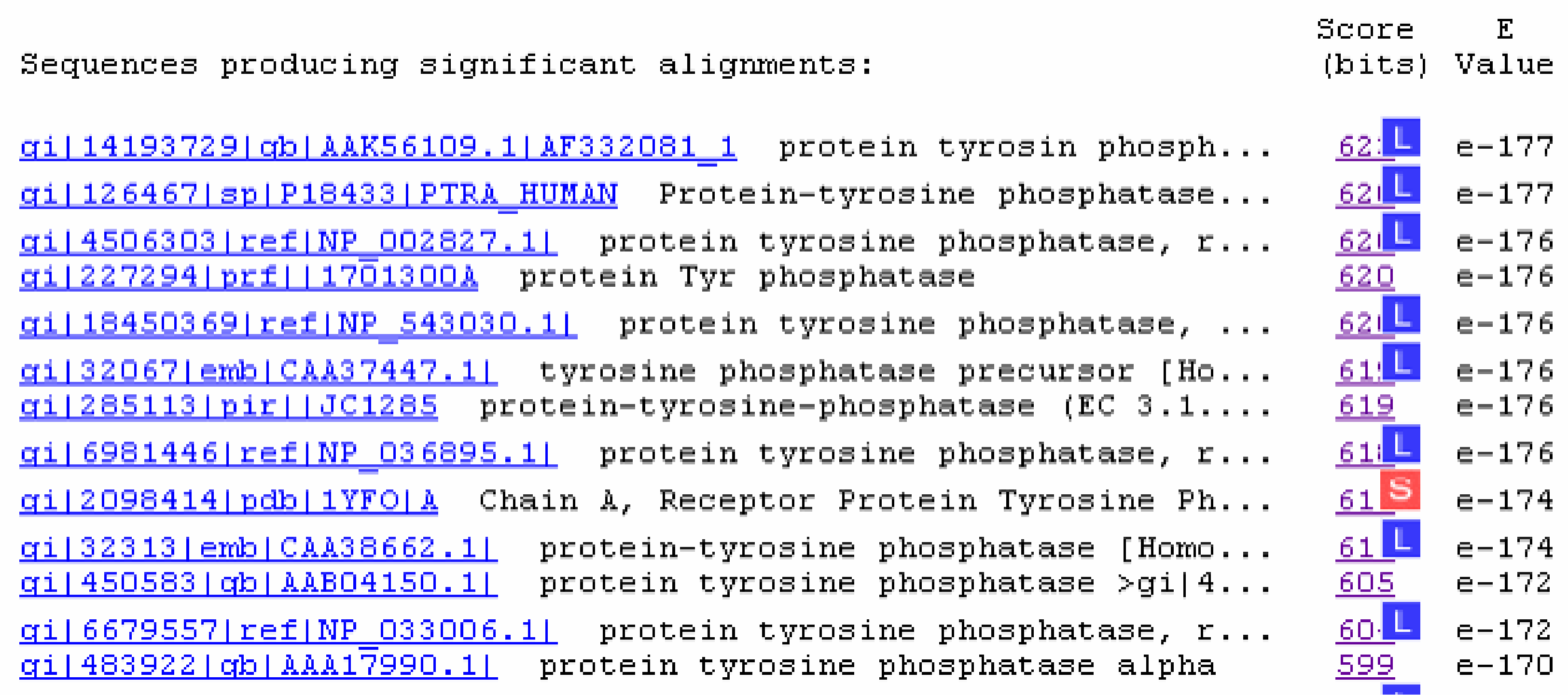
```
□ >gi|13476732|ref|NP_108301.1|   unknown protein [Mesorhizobium loti]
 gi|14027493|dbj|BAB53762.1|   unknown protein [Mesorhizobium loti]
          Length = 105

 Score =  105 bits (262), Expect = 1e-22
 Identities = 61/106 (57%), Positives = 73/106 (68%), Gaps = 1/106 (0%)

Query: 1   MKPGRLASIALAIIFLPMAVPAHAATIEITMENLVISPTEVSAKVGDTIRWVNKDVFAHT 60
            MK G L  ++         MA PA AATIE+T++ LV SP  V AKVGDTI WVN DV AHT
Sbjct: 1   MKAGALIRLSWLAALALMAAPAAAATIEVTIDKLVFSPATVEAKVGDTIEWVNNDVVAHT 60
```

good match between Amicyanin and unknown M. loti protein

Assign to *T* same function as homologs

Confirm with suitable wet experiments

Homologs by BLAST



```
                                                                   Score    E
Sequences producing significant alignments:                        (bits) Value

gi|14193729|gb|AAK56109.1|AF332081_1  protein tyrosin phosph...    62 L   e-177
gi|126467|sp|P18433|PTRA_HUMAN  Protein-tyrosine phosphatase...    62 L   e-177
gi|4506303|ref|NP_002827.1|  protein tyrosine phosphatase, r...    62 L   e-176
gi|227294|prf||1701300A  protein Tyr phosphatase                   620    e-176
gi|18450369|ref|NP_543030.1|  protein tyrosine phosphatase, ...    62 L   e-176
gi|32067|emb|CAA37447.1|  tyrosine phosphatase precursor [Ho...    61 L   e-176
gi|285113|pir||JC1285  protein-tyrosine-phosphatase (EC 3.1....    619    e-176
gi|6981446|ref|NP_036895.1|  protein tyrosine phosphatase, r...    61 L   e-176
gi|2098414|pdb|1YFO|A  Chain A, Receptor Protein Tyrosine Ph...    61 S   e-174
gi|32313|emb|CAA38662.1|  protein-tyrosine phosphatase [Homo...    61 L   e-174
gi|450583|gb|AAB04150.1|  protein tyrosine phosphatase >gi|4...    605    e-172
gi|6679557|ref|NP_033006.1|  protein tyrosine phosphatase, r...    60 L   e-172
gi|483922|gb|AAA17990.1|  protein tyrosine phosphatase alpha       599    e-170
```

- **Thus our example sequence could be a protein tyrosine phosphatase α (PTPα)**

Seq Similarity: Caveats

- **Ensure that the effect of database size and other biases has been accounted for**

- **Ensure that the function of the homology is not derived via invalid “transitive assignment’’**

- **Ensure that the target sequence has all the key features associated with the function, e.g., active site and/or domain**

Law of Large Numbers

- **Suppose you are in a room with 365 other people**
- **Q: What is the prob that a specific person in the room has the same birthday as you?**
- **A: 1/365 = 0.3%**
- **Q: What is the prob that there is a person in the room having the same birthday as you?**
- **A: 1 – $(364/365)^{365}$ = 63%**
- **Q: What is the prob that there are two persons in the room having the same birthday?**
- **A: 100%**

Interpretation of P-value

- Seq. comparison progs, e.g. BLAST, often associate a P-value to each hit

- P-value is interpreted as prob that a random seq has an equally good alignment

- Suppose the P-value of an alignment is 10^{-6}

- If database has 10^{7} seqs, then you expect $10^{7} * 10^{-6} = 10$ seqs in it that give an equally good alignment

$\Rightarrow$ Need to correct for database size if your seq comparison prog does not do that!

Lightning Does Strike Twice!

- **Roy Sullivan, a former park ranger from Virgina, was struck by lightning 7 times**
  - 1942 (lost big-toe nail)
  - 1969 (lost eyebrows)
  - 1970 (left shoulder seared)
  - 1972 (hair set on fire)
  - 1973 (hair set on fire & legs seared)
  - 1976 (ankle injured)
  - 1977 (chest & stomach burned)
- **September 1983, he committed suicide**

Cartoon: Ron Hipschman
Data: David Hand

Effect of Seq Compositional Bias

- One fourth of all residues in protein seqs occur in regions with biased amino acid composition
- Alignments of two such regions achieves high score purely due to segment composition

⇒ While it is worth noting that two proteins contain similar low complexity regions, they are best excluded when constructing alignments

- E.g., by default, BLAST employs the SEG algo to filter low complexity regions from proteins before executing a search

Source: NCBI

Seq Similarity: Caveats

- Ensure that the effect of database size and other biases has been accounted for

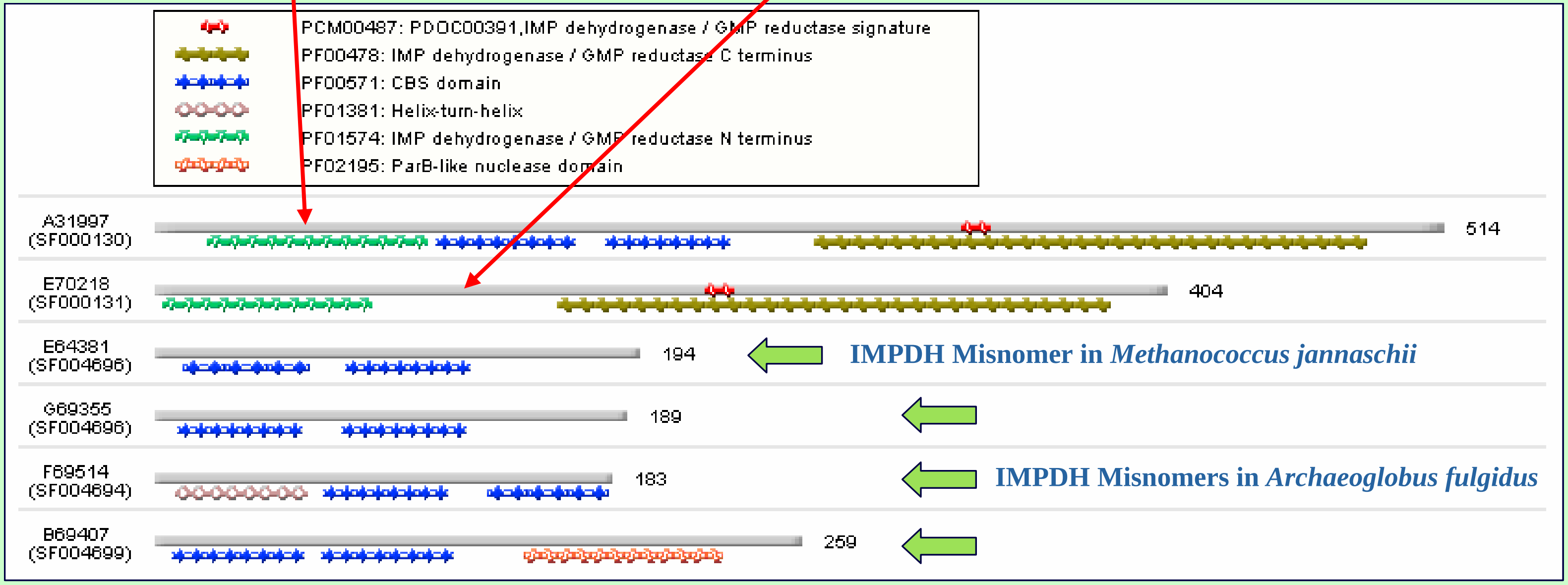
- Ensure that the function of the homology is not derived via invalid “transitive assignment’’

- **Ensure that the target sequence has all the key features associated with the function, e.g., active site and/or domain**

Emerging Pattern

Typical IMPDH

Functional IMPDH w/o CBS


PCM00487: PDOC00391,IMP dehydrogenase / GMP reductase signature
PF00478: IMP dehydrogenase / GMP reductase C terminus
PF00571: CBS domain
PF01381: Helix-turn-helix
PF01574: IMP dehydrogenase / GMP reductase N terminus
PF02195: ParB-like nuclease domain
A31997 (SF000130)
514
E70218 (SF000131)
404
E64381 (SF004696)
194
G69355 (SF004696)
189
F69514 (SF004694)
183
B69407 (SF004699)
259

IMPDH Misnomer in *Methanococcus jannaschii*

IMPDH Misnomers in *Archaeoglobus fulgidus*

- **Most IMPDHs have 2 IMPDH and 2 CBS domains**
- **Some IMPDH (E70218) lacks CBS domains**

⇒ Alignment must preserve IMPDH domain to infer IMPDH

Effect of New Approach to CS2220

- **2006 was first year of implementation**
- **9 students took the module**
  - 2 clear As
  - 2 clear C/Ds
  - 5 clear Bs
- **~50% success rate in attracting students to do more bioinformatics?**
  - 2 A students and 2 B students subsequently chose bioinformatics for individual research proj

Any Question?