(*24*, *25*). These latter RNAs may serve alternate regulatory or structural roles and await detailed characterization.

References and Notes

1. D. W. Selinger *et al.*, *Nat. Biotechnol.* **18**, 1262 (2000).
2. B. Tjaden *et al.*, *Nucleic Acids Res.* **30**, 3732 (2002).
3. K. Yamada *et al.*, *Science* **302**, 842 (2003).
4. P. Kapranov *et al.*, *Science* **296**, 916 (2002).
5. J. L. Rinn *et al.*, *Genes Dev.* **17**, 529 (2003).
6. E. F. Nuwaysir *et al.*, *Genome Res.* **12**, 1749 (2002).
7. T. J. Albert *et al.*, *Nucleic Acids Res.* **31**, e35 (2003).
8. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
9. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
10. Materials and methods are available as supporting material on *Science* Online. Additional information can be found at http://transcriptome.gersteinlab.org. Experimental data and associated microarray designs are available in the NCBI Gene Expression Omnibus (GEO) under series GSE1904, sample records GSM34073 to GSM34213, and platform records GPL1539 to GPL1673.
11. P. Bertone *et al.*, data not shown.
12. K. D. Pruitt *et al.*, *Trends Genet.* **16**, 44 (2000).
13. T. Hubbard *et al.*, *Nucleic Acids Res.* **30**, 38 (2002).
14. E. Birney *et al.*, *Genome Res.* **14**, 925 (2004).
15. Each probe is assigned a value of 1 if its fluorescence intensity is greater than the median intensity of all probes on the array, and 0 otherwise. For a given gene, the expected count of 1's within annotated exons follows a binomial distribution; an unusually high count of 1's therefore yields low *P* values (sign test). Genes having *P* values < 0.05 were regarded as demonstrating positive hybridization.
16. C. Burge, S. Karlin, *J. Mol. Biol.* **268**, 78 (1997).
17. D. L. Wheeler *et al.*, *Nucleic Acids Res.* **31**, 28 (2003).
18. Polyadenylation signals are required to appear downstream of the 15th nucleotide of the 3′ oligonucleotide in the transcribed region. An additional 51 ($46 + 5$) downstream nucleotides are included in the calculation to ensure full coverage of the sequence.
19. S. F. Altshul *et al.*, *J. Mol. Biol.* **215**, 403 (1990).
20. P. M. Harrison *et al.*, *Genome Res.* **12**, 272 (2002).
21. Z. Zhang *et al.*, *Genome Res.* **13**, 2541 (2003).
22. P. G. Buckley *et al.*, *Hum. Mol. Genet.* **11**, 3221 (2002).
23. A. S. Ishkanian *et al.*, *Nat. Genet.* **36**, 299 (2004).
24. J. S. Mattick, *Bioessays* **25**, 930 (2003).
25. D. Kampa *et al.*, *Genome Res.* **14**, 331 (2004).
26. This work was supported by NIH grant P50 HG02357.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1103388/DC1
Materials and Methods
Microarray hybridization protocols
DNA sequences of transcriptionally active regions
Fig. S1

29 July 2004; accepted 3 November 2004
Published online 11 November 2004;
10.1126/science.1103388
Include this information when citing this paper.

Use of Logic Relationships to Decipher Protein Network Organization

Peter M. Bowers,[1,2] Shawn J. Cokus,[3]
David Eisenberg,[1,2] Todd O. Yeates[2,4]*

A major focus of genome research is to decipher the networks of molecular interactions that underlie cellular function. We describe a computational approach for identifying detailed relationships between proteins on the basis of genomic data. Logic analysis of phylogenetic profiles identifies triplets of proteins whose presence or absence obey certain logic relationships. For example, protein C may be present in a genome only if proteins A and B are both present. The method reveals many previously unidentified higher order relationships. These relationships illustrate the complexities that arise in cellular networks because of branching and alternate pathways, and they also facilitate assignment of cellular functions to uncharacterized proteins.

The sequencing of multiple genomes from diverse species has tremendous potential to impact our understanding of biology, both by providing a census of all proteins and by enabling subsequent analysis of their functions (*1*–*6*). Various patterns across multiple complete genomes have been used to infer biological interactions and functional linkages between proteins (*6*–*14*). These include observations of two distinct proteins from one organism being genetically fused into a single protein in another organism (*13*, *14*) and the tendency of two proteins to occur in chromosomal proximity across multiple organisms (*12*, *15*). When a sufficiently large number of genomes were fully sequenced, it became possible with the phylogenetic profile approach (*11*, *16*, *17*) to detect functional relationships between proteins exhibiting statistically similar patterns of presence or absence. Because sequenced genomes allow us to catalog all of the proteins encoded in each organism, we can determine the pattern describing a protein's presence or absence by searching for its homologs across *N* organisms, the result of which is an *N*-dimensional vector of ones (present) and zeros (not present) referred to as its phylogenetic profile.

Original implementations of the phylogenetic profile method sought to infer "links" between pairs of proteins with similar profiles (*11*). A subsequent variation on that idea linked proteins if their profiles represented the negation of each other (*18*, *19*). These ideas are consistent with the simplest notion of how two proteins might be related in a cell, with the presence of one protein implying the presence or absence of another. Such simple patterns might be expected when two proteins are required to form a structural complex or when two proteins carry out sequential steps in an unbranched metabolic pathway. However, such simple relationships cannot adequately describe the full complexity of cellular networks that involve branching, parallel, and alternate pathways.

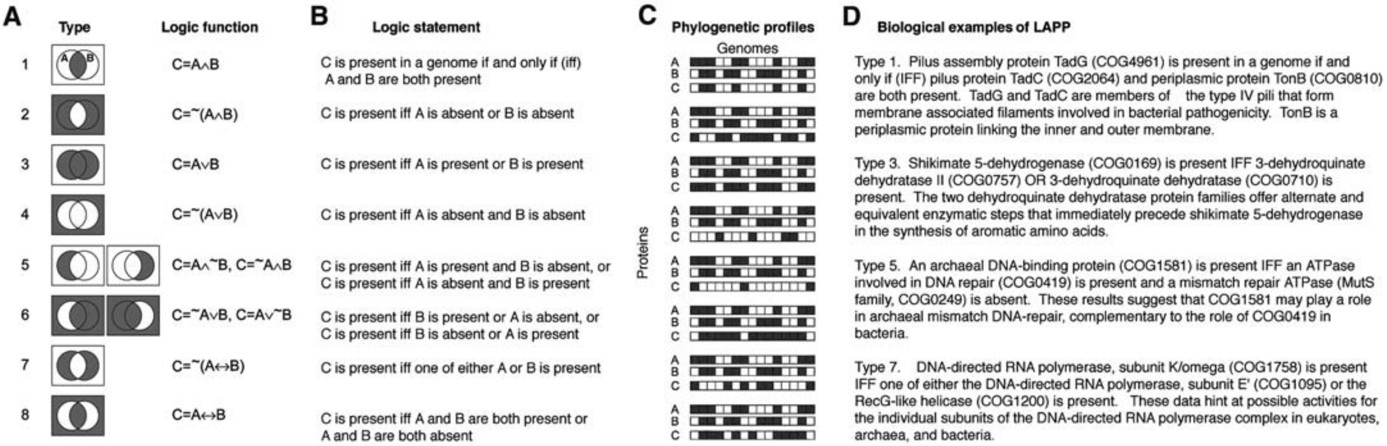
The observed complexity of cellular networks leads one to expect the existence of higher order logic relationships involving a pattern of presence or absence of multiple proteins. Furthermore, evolutionary divergence, convergence, and horizontal transfer events lead us to expect relationships between multiple gene families that are more complex than can be described by pairwise phylogenetic similarity. Analysis of cellular pathways and networks in terms of logic relations has attracted recent interest (*20*, *21*), and the growing number of sequenced genomes now makes it possible to search for logic relations.

Here, we perform a complete analysis of the logic relations possible between triplets of phylogenetic profiles and demonstrate the power of the resulting logic analysis of phylogenetic profiles (LAPP) in illuminating relationships among multiple proteins and inferring the coarse function of large numbers of uncharacterized protein families. There are eight possible logic relationships combining two phylogenetic profiles to match a third profile (Fig. 1A). For instance, protein C might be present if and only if proteins A and B are both present (denoted here as a type 1 logic relationship), from which we would infer that the function of protein C is necessary only when the functions of proteins A and B are both present. Alternatively, gene C may be present if and only if either A or B is present (a type 7 logic relationship), which is seen when different organisms use two different protein families in combination with a common third protein to accomplish some task (for example, a combination of A and C or B and C). Several of the eight possible logic relationships can be intuitively understood to describe commonly observed biological scenarios, whereas a few of the logic relationships are not easily related to real biological situations.

To identify protein triplets that exhibit the logic relationships described in Fig. 1, we first created a set of binary-valued vectors describing the presence or absence of

[1]Howard Hughes Medical Institute, [2]Institute for Genomics and Proteomics, [3]Department of Mathematics, [4]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA.

*To whom correspondence should be addressed. E-mail: yeates@mbi.ucla.edu



Type	Logic function	Logic statement
1	C=A∧B	C is present in a genome if and only if (iff) A and B are both present
2	C=~(A∧B)	C is present iff A is absent or B is absent
3	C=A∨B	C is present iff A is present or B is present
4	C=~(A∨B)	C is present iff A is absent and B is absent
5	C=A∧~B, C=~A∧B	C is present iff A is present and B is absent, or C is present iff A is absent and B is present
6	C=~A∨B, C=A∨~B	C is present iff B is present or A is absent, or C is present iff B is absent or A is present
7	C=~(A↔B)	C is present iff one of either A or B is present
8	C=A↔B	C is present iff A and B are both present or A and B are both absent

Biological examples of LAPP

Type 1. Pilus assembly protein TadG (COG4961) is present in a genome if and only if (IFF) pilus protein TadC (COG2064) and periplasmic protein TonB (COG0810) are both present. TadG and TadC are members of the type IV pili that form membrane associated filaments involved in bacterial pathogenicity. TonB is a periplasmic protein linking the inner and outer membrane.

Type 3. Shikimate 5-dehydrogenase (COG0169) is present IFF 3-dehydroquinate dehydratase II (COG0757) OR 3-dehydroquinate dehydratase (COG0710) is present. The two dehydroquinate dehydratase protein families offer alternate and equivalent enzymatic steps that immediately precede shikimate 5-dehydrogenase in the synthesis of aromatic amino acids.

Type 5. An archaeal DNA-binding protein (COG1581) is present IFF an ATPase involved in DNA repair (COG0419) is present and a mismatch repair ATPase (MutS family, COG0249) is absent. These results suggest that COG1581 may play a role in archaeal mismatch DNA-repair, complementary to the role of COG0419 in bacteria.

Type 7. DNA-directed RNA polymerase, subunit K/omega (COG1758) is present IFF one of either the DNA-directed RNA polymerase, subunit E' (COG1095) or the RecG-like helicase (COG1200) is present. These data hint at possible activities for the individual subunits of the DNA-directed RNA polymerase complex in eukaryotes, archaea, and bacteria.

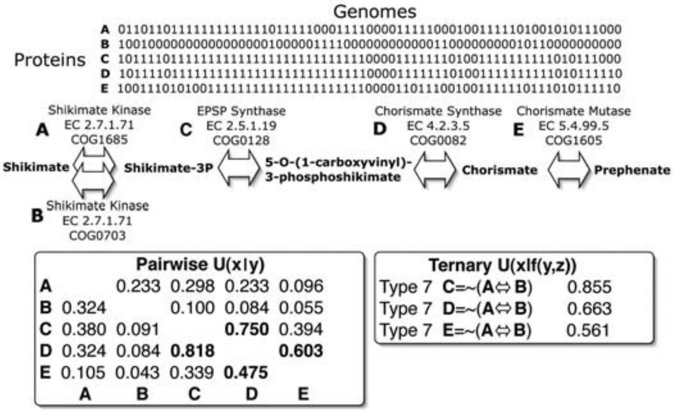
Fig. 1. Detection of pathway relationships among proteins based on a logic analysis of phylogenetic profiles (LAPP). (**A**) Venn diagrams and associated logic statements illustrate the eight distinct kinds of logic functions that describe the possible dependence of the presence of C on the presence of A and B, jointly. Logic functions are grouped together if they are related by a simple exchange of proteins A and B. The symbols ∧, ∨, ~, and ↔ indicate "logical AND," "logical OR," "logical negation," and "logical equality," respectively. (**B**) The meaning of each logic relationship is described in a single text sentence, (**C**) hypothetical phylogenetic profiles are used to illustrate the eight possible logic functions, and (**D**) for the four most commonly observed logic types, real biological examples are given that illustrate the ternary relationships identified from actual phylogenetic profiles.

each of the known protein families across 67 fully sequenced organisms. Specifically, the complete set of proteins was categorized into 4873 distinct families known as clusters of orthologous groups (COGs) (*22*, *23*). Next, we systematically examined all triplet combinations of the profiles and rank-ordered them according to how well the logical combination $f(a,b)$ of two profiles predicted a third profile, c. Further, we also required that neither profile a nor b alone was predictive of c. We calculated uncertainty coefficients for $U(c|a)$, $U(c|b)$, and the logically combined profile $U(c|f(a,b))$, where

$$U(x|y) = [H(x) + H(y) - H(x,y)]/H(x)$$

and H refers to the entropy of the individual or joint distributions (*24*). The value of U can range between 1.0, where x is a deterministic function of y, and 0.0, where x is completely independent of y. We selected triplets whose individual pairwise uncertainty scores described protein profile c poorly [$U(c|a) < 0.3$ and $U(c|b) < 0.3$] but whose logically combined profile [$U(c|f(a,b)) > 0.6$] described c well.

A hypothetical set of profiles can illustrate the approach (Fig. 1C). Under a type 3 logic relationship, protein C is present whenever protein A, protein B, or both are present. The pairwise comparisons of profiles (AC, BC, and AB) each yield limited information, whereas a phylogenetic profile logically combining proteins A and B matches the phylogenetic distribution of protein C exactly and has a triplet uncertainty score of $U = 0.48$. In contrast, a triplet containing hypothetical randomized profiles with the same number of protein homologs as in the previous example has a triplet uncertainty coefficient of $U = 0.03$ and does not correspond closely to any of the eight logic types.


Genomes
Proteins
A 0110110111111111111111101111100011110000111110001001111101001010111000
B 1001000000000000000000010000011111000000000000001100000000001011000000000
C 1011110111111111111111111111111111100001111111010011111111111010111000
D 1011110111111111111111111111111111100001111111010011111111111010111110
E 1001110101001111111111111111111111100001101110010011111110111010111110
Shikimate Kinase
A
EC 2.7.1.71
COG1685
B
Shikimate Kinase
EC 2.7.1.71
COG0703
EPSP Synthase
C
EC 2.5.1.19
COG0128
Chorismate Synthase
D
EC 4.2.3.5
COG0082
Chorismate Mutase
E
EC 5.4.99.5
COG1605
Shikimate
Shikimate-3P
5-O-(1-carboxyvinyl)-3-phosphoshikimate
Chorismate
Prephenate

Pairwise U(x|y)

A		0.233	0.298	0.233	0.096
B	0.324		0.100	0.084	0.055
C	0.380	0.091		**0.750**	0.394
D	0.324	0.084	**0.818**		**0.603**
E	0.105	0.043	0.339	**0.475**	
	A	B	C	D	E

Ternary U(x|f(y,z))

Type 7	**C=~(A⇔B)**	0.855
Type 7	**D=~(A⇔B)**	0.663
Type 7	**E=~(A⇔B)**	0.561

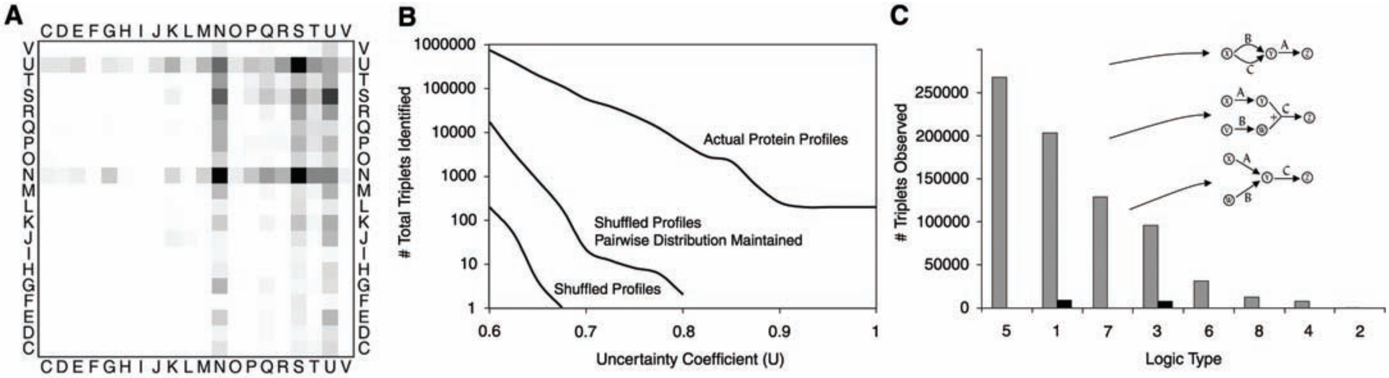
Fig. 2. Three logic examples from the aromatic amino acid synthesis pathway, obtained as high-scoring ternary relationships in an analysis of all possible 62 billion protein triplets. In this example, a calculation based on traditional pairwise phylogenetic profile analysis links only the terminal enzymes in the pathway (proteins C and D and E). The triplet and pairwise uncertainty coefficients, U, highlight the additional associations observed with ternary relationships, A and B with C, D, and E.

Logic analysis of phylogenetic profiles yields thousands of computed relationships among protein families that cannot be detected by traditional pairwise phylogenetic analysis, enabling a more intricate description of predicted relationships (Fig. 2 and fig. S1). The synthesis of aromatic amino acids proceeds through the shikimate pathway. A logic analysis of five participating proteins shows that shikimate can be converted to the end product prephenate by one of two possible routes, leading to a type 7 logic relationship. When either one shikimate kinase protein family (protein A, COG1685) or an alternate shikimate kinase protein family (protein B, COG0703) is present in an organism, then excitatory postsynaptic potential (EPSP) synthase must also be present (protein C, COG0128) ($U = 0.85$) to carry out the subsequent enzymatic step. The same type 7 logic relationship is also observed between alternate shikimate kinase enzymes and the successive chorismate synthase (protein D, COG0082) and chorismate mutase (protein E, COG1605) enzymatic steps of the pathway. The ordering of the metabolic steps that follow shikimate kinase is predicted by the value of successive U coefficients, where EPSP synthase (second step, $U = 0.85$) is most strongly linked to shikimate kinase, followed directly by the chorismate synthase (third step, $U = 0.66$) and lastly by chorismate mutase (fourth step, $U = 0.56$). We can conclude that organisms synthesize chorismate and prephenate from shikimate with the use of only one of two possible alternate routes: pathways consisting of either ordered enzymes A-C-D-E or enzymes B-C-D-E.

Our LAPP recovers 750,000 previously unknown relationships among protein families ($(U(c|(f(a,b))) > 0.60$; $U(c|b) < 0.30$; $U(c|a) < 0.30$), whose validity can be assessed by comparing known annotations of the linked proteins (tables S2 to S5). The ability to recover links between proteins annotated as belonging to a major functional category has been used widely to corroborate computational inferences of protein interactions (*4*, *5*). We assessed the accuracy of the logic relationships obtained for the triplet profiles by using the metrics and threshold values detailed above, where each protein family is annotated as belonging to one or more of the 20 COG major functional categories. Figure 3A shows a section taken from a three-dimensional histogram that describes the frequency of observed logic relationships in which protein A of the triplet is annotated as belonging to the COG functional category N, cell motility. One of the most frequently observed triplet relationships in this section relates three proteins belonging to the cell motility category, confirmation that the triplet associations link proteins closely related in function. Other triplets involve two proteins from the motility category and a third protein of another COG category, producing recognizable horizontal and vertical bands in the histogram. For instance, the category combinations NNU (COG category U, intracellular trafficking and secretion) and NNS (COG category S, unknown function) are also plentiful. Connections between these categories make intuitive sense and facilitate placement of unannotated proteins within the context of specific cellular networks of interacting proteins.

The LAPP method leads to a set of statistically significant ternary relationships (Fig. 3B and fig. S2) that are distinct from and more numerous than the relationships that can be inferred by using traditional pairwise analysis. A matrix of randomized phylogenetic profiles, containing the same individual and pairwise distributions as the native profiles, was used to assess the probability of observing a given uncertainty coefficient score by chance. Triplets with $U > 0.60$ are observed from the unshuffled vectors $\sim10^2$ times more frequently than from shuffled profiles and $\sim10^4$ more frequently when $U > 0.80$. A P value for each triplet relationship can be calculated by enumerating all possible values of U that could be obtained from shuffled profiles while maintaining the individual and pairwise distributions, where P is equal to the number of trials that exceed the observed value of U divided by the total number of trials. More than 98% of the identified triplets ($U > 0.6$) have $P < 0.05$, and more than 75% of the identified triplets have $P < 0.005$. Lastly, the eight distinct logic types occur with widely varying frequencies within the set of significant ternary relationships (Fig. 3C), a trend consistent with our understanding of evolution and biological relationships. Logic types 1, 3, 5, and 7 are observed frequently in the biological data, whereas logic types 2, 4, and 8 are more difficult to relate to simple cellular logic and are observed only rarely.

The 50 most significant computed ternary relationships from Fig. 3A are shown in network form (Fig. 3D). The proteins linked include secreted virulence factors, adhesin proteins necessary for bacterial pathogenesis,



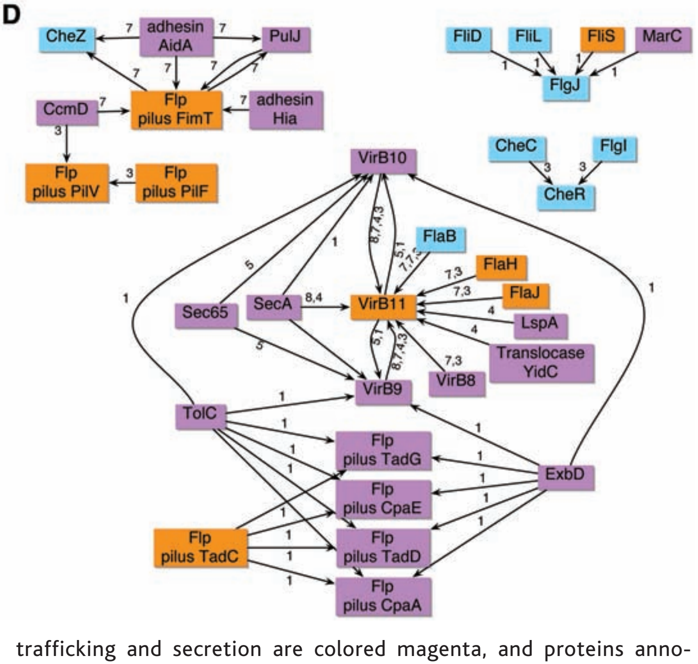


Fig. 3. A benchmark analysis of the ability of LAPP to identify functionally related proteins. (**A**) A section from a three-dimensional histogram describing the prevalence of ternary relationships among high-scoring triplets. The histogram is for logic function type 3 and covers triplets of proteins A, B, and C whose COG major functional categories (table S3) are described by N, *x* axis, and *y* axis, where N is cell motility. Because protein families are not uniformly distributed across COG categories, *Z* scores are plotted to facilitate comparison of histogram bin counts. The mean μ and variance σ^2 for each bin was calculated for a distribution of 750,000 triplets with randomly selected protein families, and an observed bin count, *n*, was transformed by $Z = (n - \mu)/\sigma$; the gray scale is linear, with white corresponding to $Z = 0.0$ and black to $Z = 75$. (**B**) A plot of the cumulative number of protein triplets recovered at an uncertainty coefficient score greater than a given threshold. LAPP analysis of a randomized matrix, containing shuffled profiles that preserved the overall individual and pairwise distributions (fig. S2), reveals only ~20,000 triplets with high coefficient scores. In contrast, we detect 750,000 triplets from an analysis of the original, unshuffled biological protein profiles. (**C**) A histogram showing the number of identified triplets ($U > 0.6$) for each of the eight logic function types for randomized (black) and real (gray) phylogenetic profiles. Diagrams illustrate one possible pathway arrangement consistent with that type of logic. Some diagrams describe potential pathways combined across multiple organisms, but the three proteins of interest may not always occur together in any single given genome. (**D**) An illustration of the 50 highest scoring relationships ($U > 0.75$) involving proteins from the cell motility and intracellular trafficking and secretion functional categories. Cell motility proteins are colored light blue, intracellular trafficking and secretion are colored magenta, and proteins annotated as both are colored in orange. Edges are shown between proteins A-C and B-C of each logic triplet, with each edge labeled according to the logic function type used to associate the proteins' families.

chemotaxis proteins, and translocase proteins. The network contains previously unknown interactions that suggest mechanisms connecting bacterial pathogenesis and chemotaxis. For instance, CheZ, a chemotaxis dephosphorylase that regulates cell motility, is linked to the surface receptor and virulence factors adhesin AidA and Flp pilus-associated FimT.

The new higher order protein associations detected by LAPP provide a framework for understanding the complex logical dependencies that relate proteins to one another in the cell. They may also be useful in modeling and engineering biological systems, generating biological hypotheses for experimentation, and investigating additional protein properties. It is likely that the logic relationships between proteins in the cell extend beyond ternary relationships to include much larger sets of proteins. We anticipate that the ideas underlying the logical analysis of phylogenetic profiles can be extended to the investigation of other kinds of genomic data, such as gene expression, nucleotide polymorphism, and phenotype data.

References and Notes

1. S. Li *et al.*, *Science* **303**, 540 (2004); published online 2 January 2004 (10.1126/science.1091403).
2. M. Strong, P. Mallick, M. Pellegrini, M. J. Thompson, D. Eisenberg, *Genome Biol.* **4**, R59 (2003).
3. L. Giot *et al.*, *Science* **302**, 1727 (2003); published online 6 November 2003 (10.1126/science.1090289).
4. P. M. Bowers *et al.*, *Genome Biol.* **5**, R35 (2004).
5. C. von Mering *et al.*, *Nucleic Acids Res.* **31**, 258 (2003).
6. A. H. Y. Tong *et al.*, *Science* **303**, 808 (2004).
7. Y. Ho *et al.*, *Nature* **415**, 180 (2002).
8. A. C. Gavin *et al.*, *Nature* **415**, 141 (2002).
9. T. Ito *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569 (2001).
10. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
11. M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, T. O. Yeates, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4285 (1999).
12. R. Overbeek, M. Fonstein, M. D'Souza, G. D. Pusch, N. Maltsev, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2896 (1999).
13. E. M. Marcotte *et al.*, *Science* **285**, 751 (1999).
14. A. J. Enright, I. Iliopoulos, N. C. Kyrpides, C. A. Ouzounis, *Nature* **402**, 86 (1999).
15. M. D. Ermolaeva, O. White, S. L. Salzberg, *Nucleic Acids Res.* **29**, 1216 (2001).
16. M. Pellegrini, M. Thompson, J. Fierro, P. Bowers, *J. Cell. Biochem. Suppl.* **37**, 106 (2001).
17. J. Wu, S. Kasif, C. DeLisi, *Bioinformatics* **19**, 1524 (2003).
18. E. Morett *et al.*, *Nat. Biotechnol.* **21**, 790 (2003).
19. S. V. Date, E. M. Marcotte, *Nat. Biotechnol.* **21**, 1055 (2003).
20. R. Milo *et al.*, *Science* **298**, 824 (2002).
21. R. Milo *et al.*, *Science* **303**, 1538 (2004).
22. R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997).
23. R. L. Tatusov *et al.*, *BMC Bioinformatics* **4**, 41 (2003).
24. H. Theil, *Statistical Decomposition Analysis with Applications in the Social and Administrative Sciences*, vol. 14 of *Studies in Mathematical and Managerial Economics* (North-Holland, Amsterdam, 1972), pp. xvi, 337.

25. This work was supported by the U.S. Department of Energy and the Howard Hughes Medical Institute.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5705/2246/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 and S2

28 July 2004; accepted 22 November 2004
10.1126/science.1103330

Reproductive Effort, Molting Latitude, and Feather Color in a Migratory Songbird

D. Ryan Norris,[1,4]* Peter P. Marra,[3] Robert Montgomerie,[1] T. Kurt Kyser,[2] Laurene M. Ratcliffe[1]

Toward the end of the breeding season, migratory songbirds face crucial tradeoffs between the timing of reproduction, molt, and migration. Using stable hydrogen isotopes, we show that male American redstarts investing in high levels of reproduction late in the season adopt a unique strategy of combining molt and migration. Tail feathers molted during migration also reflect less orange-red light, indicating reduced carotenoid concentration. Thus, we show how reproduction in a migratory animal can influence both life history strategies (location of molt) and social signals (feather color) during subsequent periods of the annual cycle.

Each year, toward the end of the temperate breeding season, billions of songbirds face crucial energetic tradeoffs between the costs of reproduction, the replacement of feathers (molt), and the hazards of long-distance migration to the tropics (*1*). To date, our inability to track individual birds moving between their breeding and wintering grounds has made studying the interaction between these events virtually impossible. Using stable hydrogen isotopes and reflectance spectrometry, we investigate how reproduction affects both molting latitude and the color of molted feathers in an 8-g neotropical-nearctic migratory songbird, the American redstart (*Setophaga ruticilla*).

Redstarts (Fig. 1A) are socially monogamous, single-brooded passerine birds that provide biparental care to young for 2 to 3 weeks after the young leave the nest (*2*). Individuals breed in the deciduous forests of temperate North America and winter in the Caribbean and Middle America. From 2001 to 2004, we sampled tail feathers from individually marked males at a breeding site in Ontario, Canada (44°34′ N, 76°19′ W). These males were known to have bred at the same location the previous year (*3*). In eastern North America, stable hydrogen isotope (δD) values in precipitation follow a strong latitudinal gradient where low (more negative) values correspond to higher latitudes (Fig. 1B) (*4*). δD signatures in precipitation are transferred through food webs to higher-order consumers, including birds (*5*). Because feathers are metabolically inert after growth, δD values sampled from feathers in a given breeding season indicate the molting latitude from the previous autumn.

[1]Department of Biology, [2]Department of Geological Sciences and Engineering, Queen's University, Kingston, Ontario K7L 3N6, Canada. [3]Smithsonian Environmental Research Center, Post Office Box 28, 647 Contees Wharf Road, Edgewater, MD 21037, USA. [4]Centre for Applied Conservation Research, Forest Sciences Center, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

*To whom correspondence should be addressed. E-mail: ryann@biology.queensu.ca

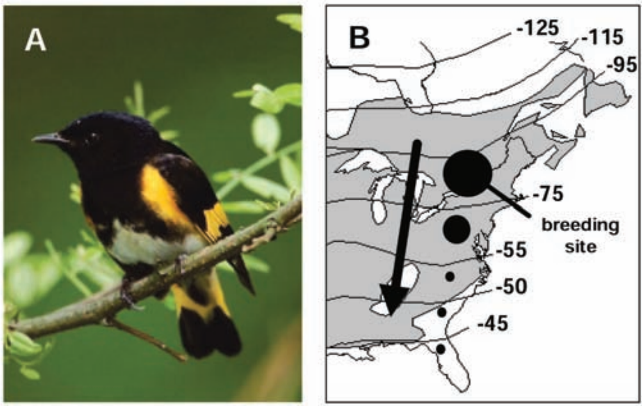


Fig. 1. (**A**) Adult male American redstart after complete autumn molt. [Photograph by Robert Royse] (**B**) Distribution of post-breeding molt locations determined from δD values of tail feathers (n = 30). Contour lines indicate expected δD values throughout eastern North America (*4*). The eastern portion of the breeding range is shaded light gray (*2*). The size of the circles represents the frequency distribution of molt locations: large (near breeding grounds), n = 18 individuals; medium, n = 9; small, n = 1. The arrow shows the most likely fall migration route based on band-recapture data (*16*).