

and Leu<sup>402/a</sup>). Moreover, the side chain of Leu<sup>362/a</sup> is buried within the hydrophobic cleft formed by MurNAc(anhydro) from TCT, and Ala<sup>368/x</sup>, Val<sup>383/x</sup>, and Phe<sup>387/x</sup> (Fig. 4A). Most of the interface residues in LCa show an induced-fit conformational change (Fig. 4B). Likewise, the orientations of many side chains of the LCx interface residues may also be induced as a result of TCT docking. The indole ring of Trp<sup>392/x</sup> is stacked against the indole ring of Trp<sup>394/x</sup>, which in turn stacks against the elongated side chain of DAP (Fig. 3); the aromatic ring of Phe<sup>387/x</sup> is in intimate van der Waals contact with MurNAc. Without bound TCT, both Phe<sup>387/x</sup> and Trp<sup>392/x</sup> would not have their side chains oriented properly for LCa interaction. This notion is favored by previous observation that a lactyl-tetrapeptide still retains some stimulatory activity despite the lack of a disaccharide moiety (3, 5); presumably, the stem peptide alone can bind to the LCx docking groove and induce properly oriented Trp<sup>392/x</sup> through stacking interactions to allow heterodimerization.

In conclusion, the structure of the LCa-TCT-LCx complex has revealed the molecular interaction between a naturally occurring PG subunit and its PRR complex. The complex structure shows that TCT binds to the docking groove of LCx primarily through its elongated stem peptide and that the bound TCT conformation, which allows its disaccharide GlcNAc-MurNAc(anhydro) to be presented for LCa interaction, is further stabilized by extensive hydrogen bonding. DAP,

the specificity determinant of TCT, engages a key electrostatic interaction with Arg<sup>413/x</sup>. Stacking interactions with the DAP side chain and the MurNAc ring induce proper orientations of side chains of Phe<sup>387/x</sup> and Trp<sup>392/x</sup>, which contribute part of the TCT-LCx interface for LCa binding by induced fit. The present complex structure represents the receptor activation step where a monomeric PG ligand brings PGRP-LCa and -LCx receptors together to allow cytoplasmic domains to interact for activation of the immune deficiency pathway. By contrast, activation of this pathway by polymeric PG only requires PGRP-LCx, which can be brought to close proximity by binding to adjacent muropeptide subunits (9, 10). Protein-protein interaction induced by monomeric ligand as visualized by the present work is likely to serve as the general mechanism for pattern recognition for other PGRPs, such as PGRP-SA, that play a double role as both a PG sensor and a signaling molecule.

*Note added in proof:* During the revision of this paper, the structure of TCT bound to PGRP-LE was reported (19).

#### References and Notes

1. R. Medzhitov, C. A. Janeway Jr., *Cell* **91**, 295 (1997).
2. J. A. Hoffmann, *Nature* **426**, 33 (2003).
3. T. Kaneko *et al.*, *Immunity* **20**, 637 (2004).
4. F. Leulier *et al.*, *Nat. Immunol.* **4**, 478 (2003).
5. C. R. Stenbak *et al.*, *J. Immunol.* **173**, 7339 (2004).
6. T. Werner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13772 (2000).
7. D. Mengin-Lecreulx, B. Lemaître, *J. Endotoxin Res.* **11**, 105 (2005).

8. W. E. Goldman, D. G. Klapper, J. B. Baseman, *Infect. Immun.* **36**, 782 (1982).
9. C. I. Chang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10279 (2005).
10. P. Mellroth *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6455 (2005).
11. R. Guan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17168 (2004).
12. K. M. Choe, H. Lee, K. V. Anderson, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1122 (2005).
13. Materials and methods are available as supporting material on Science Online.
14. C. I. Chang *et al.*, *PLoS Biol.* **2**, E277 (2004).
15. R. Guan, Q. Wang, E. J. Sundberg, R. A. Mariuzza, *J. Mol. Biol.* **347**, 683 (2005).
16. J. B. Reiser, L. Teyton, I. A. Wilson, *J. Mol. Biol.* **340**, 909 (2004).
17. M. S. Kim, M. Byun, B. H. Oh, *Nat. Immunol.* **4**, 787 (2003).
18. C. Liu, E. Gelius, G. Liu, H. Steiner, R. Dziarski, *J. Biol. Chem.* **275**, 24490 (2000).
19. J. H. Lim *et al.*, *J. Biol. Chem.* 10.1074/jbc.M513030200 (20 January 2006).
20. We thank K. Ihara and S. Wakatsuki for sharing their beamtime for screening crystals at the Photon Factory, Japan, and for helpful advice and technical assistance. This work was supported by the Howard Hughes Medical Institute (to J.D.) and the Centre National de la Recherche Scientifique (to D.M.-L.). Coordinates and structural factors have been submitted to the Protein Data Bank (PDB) with accession number 2F2L.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/311/5768/1761/DC1  
Materials and Methods  
Tables S1 and S2  
References

28 November 2005; accepted 21 February 2006  
10.1126/science.1123056

## The Effect of Oxygen on Biochemical Networks and the Evolution of Complex Life

Jason Raymond<sup>1</sup> and Daniel Segre<sup>1,2</sup>

The evolution of oxygenic photosynthesis and ensuing oxygenation of Earth's atmosphere represent a major transition in the history of life. Although many organisms retreated to anoxic environments, others evolved to use oxygen as a high-potential redox couple while concomitantly mitigating its toxicity. To understand the changes in biochemistry and enzymology that accompanied adaptation to O<sub>2</sub>, we integrated network analysis with information on enzyme evolution to infer how oxygen availability changed the architecture of metabolic networks. Our analysis revealed the existence of four discrete groups of networks of increasing complexity, with transitions between groups being contingent on the presence of key metabolites, including molecular oxygen, which was required for transition into the largest networks.

Because the myriad biochemical mechanisms that modern organisms possess for detoxifying reactive oxygen derivatives had yet to evolve (1), the introduction of O<sub>2</sub> into an anaerobic biosphere around 2.2 billion

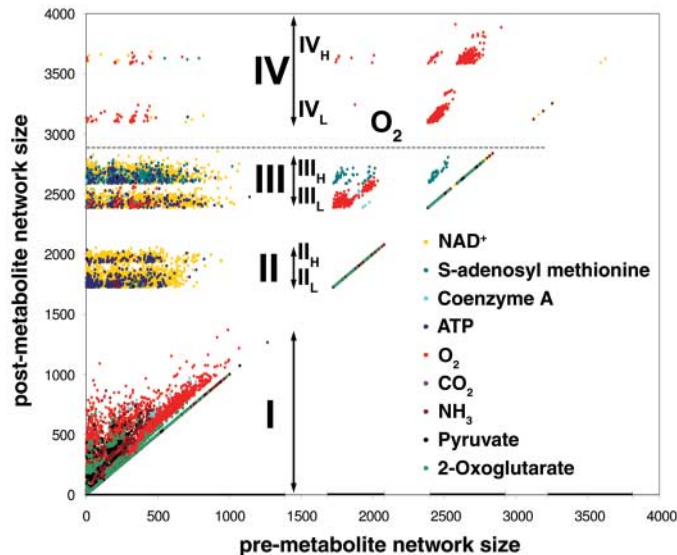
years ago must have represented a cataclysm in the history of life (2). Strong selection pressure on ancient microbes, such as the ancestors of modern cyanobacteria that were first responsible for producing O<sub>2</sub> via oxygenic photosynthesis, precipitated the development of novel enzymes and new pathways. These included adaptations that mediated the biotoxicity of O<sub>2</sub> derivatives, including superoxide and hydroxyl radicals, and countered the inherent O<sub>2</sub>/redox

sensitivity of many proteins, such as iron-sulfur cluster-containing electron shuttles (3). Recent evidence suggests that as molecular oxygen became integral to biochemical pathways, many enzymatic reactions central to anoxic metabolism were effectively replaced in aerobic organisms (4). The availability of molecular oxygen also made possible a highly exergonic respiratory chain based on O<sub>2</sub> as a terminal electron acceptor, an event that is widely held to have been closely coincident with, if not requisite for, the development of complex eukaryotic life (5, 6) and for adaptation to electron acceptor-limited terrestrial environments. The evolution of cyanobacteria and the ensuing oxidation of Earth's atmosphere left distinct, though not entirely congruent, signatures in the geological record (7–9) and provide an opportunity to understand the evolution of biochemical pathways by integrating geological clues with metabolic network modeling and phylogenetic analysis. By comparing metabolic networks attainable under oxic and anoxic conditions, we have attempted to elucidate the global metabolic reorganization resulting from this major environmental shift.

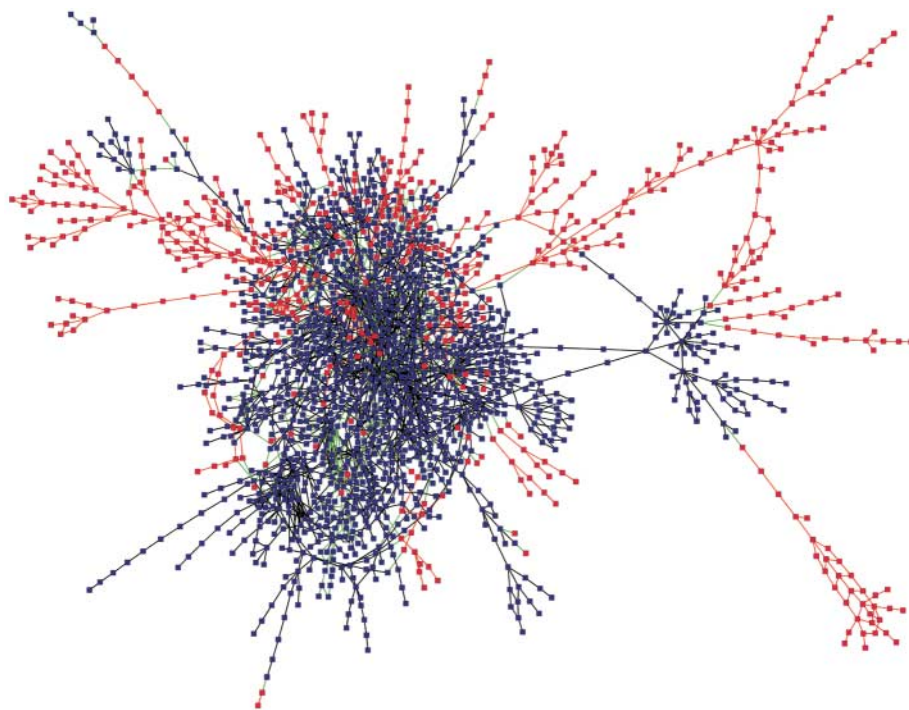
Our primary goal was to determine what effect the presence or absence of common biomolecules such as O<sub>2</sub> has on the complexity—the overall structure, including size and connectivity—of

<sup>1</sup>Microbial Systems Division, Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. <sup>2</sup>Bioinformatics Program, Department of Biology and Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA.

**Fig. 1.** The effect of various metabolites (legend at right) on the total number of reactions in ecosystem-level metabolic networks, as computed with the network expansion algorithm. Each point represents two consecutively generated networks: The first network, whose size is the abscissa value, is generated from a randomly chosen set of seed metabolites, and the second network, whose size is the ordinate value, is generated from that same seed set amended with the addition of one of the nine metabolites



shown in the legend. Points are colored based on the amended metabolite, shown in the legend. All networks occupy four broadly similar groups (bold lines and roman numerals) and subgroups (H, higher; L, lower) that result from often very different but chemically interconvertible seed sets. Only networks that include  $O_2$  as a metabolite are able to transition into group IV (dashed line), with other transitions being determined by the availability of key metabolites.



**Fig. 2.** The effect of oxygen on KEGG's metabolic "backbone" (pictured as a pruned version of the full network, with 1861 metabolites and 2652 possible reactions; see SOM for detailed description). Blue nodes and edges represent metabolites and reactions, respectively, that are present in anoxic metabolic networks. Red nodes and edges are oxic network metabolites and reactions whose presence is contingent on oxygen availability either directly or secondarily. Green edges correspond to reactions that are found only in the oxic network but use at least one anoxic metabolite, representing replacement or rewiring of anoxic pathways to take advantage of oxygen, as in thiamin and  $B_{12}$  biosynthesis.

metabolic networks. This process was carried out following a heuristic recently developed in detail by Ebenhöh *et al.* (10) and referred to as metabolic network expansion, whereby a set of pre-specified "seed" compounds are allowed to

react according to enzymatic reaction rules, for example as enumerated by the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (11). As implemented here, a reaction can occur only if all of its reactants are present in

the seed set. Once all possible reactions have been carried out, the products of those reactions then join the seed compounds, potentially allowing new reactions to occur. This process is iterated until convergence: when no new products are generated and no new reactions are possible (12) (fig. S1). The network converged upon is thereby a product of the initial seed metabolites and the KEGG reactions that are allowed [all KEGG reactions were allowed here, although additional constraints can be added in the form of selection rules (10); for example, based on the phylogenetically inferred antiquity of enzymes].

Because the KEGG database is a collection of data from across known genomes, these metabolic networks correspond not to the reactions tenable within any one organism but to the metabolic potential of the collective (and currently characterized) biosphere that can be thought of as a meta-metabolome. The initial pool of seed metabolites was determined by two procedures: random sampling of metabolites by Monte Carlo (MC) simulations and deterministic selection from putatively prebiotic compounds [see the supporting online material (SOM) for detailed information]. The current KEGG database (11, 13) encompasses 6836 reactions extending across 70 genomes and involving 5057 distinct compounds.  $O_2$  is among the most-utilized compounds, superseding biomolecules such as adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide ( $NAD^+$ ) in rank (table S1).  $O_2$  is further distinguished by the steep thermodynamic gradient favoring its reduction, resulting in oxygen being produced only by a single biological reaction: oxygenic photosynthesis. Although  $O_2$  does occur as a by-product of some biological reactions, such as that of the enzyme catalase, those reactions depend ultimately on the presence of oxygen, so that they do not result in net  $O_2$  generation. In contrast, based on current information about reaction reversibility, most compounds are produced in roughly the same number of reactions that they are consumed in ( $r^2 = 0.926$ ,  $P < 0.01$ ).

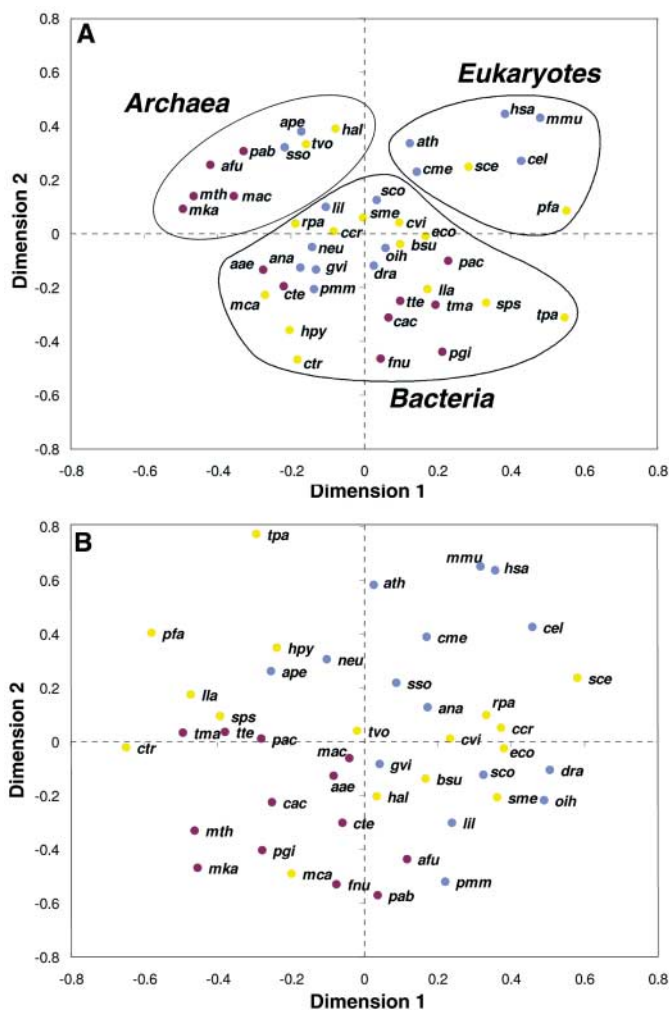
MC sampling of highly variable seed conditions and simulation of  $\sim 10^5$  networks show that all resultant metabolic networks converge to just four discrete groups of increasing size and connectivity (Fig. 1). These four groups act as basins of attraction in network space, converged on from often completely different and limitless ( $\sim 10^{16536}$  distinct seed sets) initial conditions, but sharing within each group  $>95\%$  identical reactions and metabolites. Further, the networks in smaller groups are largely nested within those in larger groups; for example, most reactions and metabolites in group II networks are a subset of those in group III and IV networks. This coalescence of growing networks into distinct groups is consistent with models of hierarchical modularity in metabolic networks (14, 15), a consequence of the asymmetric distribution of biomolecule con-

nectivity (some biomolecules are used in a disproportionately large number of reactions), the intrinsic reversibility and cyclic nature of many biochemical pathways, and gene duplication as a primary mechanism for expanding complexity (16) (see SOM for further discussion).

Networks simulated in the presence of molecular oxygen were able to transition into a group (Fig. 1, group IV) unreachable under any anoxic conditions explored by MC simulation. Networks in group IV had as many as  $10^3$  reactions more than those of the largest networks achieved in the absence of  $O_2$ . On average, 52% of these additional reactions do not explicitly use  $O_2$ , indicating the presence of pathways whose function is contingent on but does not require molecular oxygen. Among these  $O_2$ -dependent pathways are several steps associated with aerobic cobalamin synthesis, confirming that this approach successfully discovers pathway-level contingencies not detectable by previous analysis that focused on  $O_2$ -dependent enzymes as proxies for  $O_2$ -dependent pathways (4). Figure S2, A to F, indicates that the effect of oxygen on biochemical network expansion is robust to reaction database modifications, such as the characterization and addition of new reactions to KEGG (17).

Transitions between smaller groups, as well as the subsplitting of groups II to IV into “high” (H) and “low” (L) clusters (Fig. 1), are determined by the availability of biomolecules involved in the assimilation and cycling of key elements and whose essentiality may have manifested early in the evolution of metabolism (17, 18). For example, sulfur-compound availability governs the splitting in each of the three larger groups (II to IV). Elemental sulfur is required in the synthesis of a broad range of compounds, such as the amino acids methionine and cysteine and, by way of cysteine, coenzyme A (CoA). In turn, CoA is essential to organisms across all three domains of life, most notably in central carbon cycling and in biosynthesis and metabolism involving a diverse range of lipids. Including sulfur in seed networks, by way of a highly utilized “hub” compound such as sulfide or cysteine, allows group II and group III<sub>L</sub> networks to transition into group III<sub>H</sub>, as well as the group IV<sub>L</sub> → IV<sub>H</sub> transition in oxic networks. Other highly connected compounds, particularly those central to cycling fundamental elements such as carbon, nitrogen, and phosphorus, appear to be responsible for the discrete clusters of network sizes, as shown for example in Fig. 1 and as recently elucidated by Ebenhöf *et al.* (19).

Genomic information provides insight into how biomolecule availability and use shape enzyme content in diverse organisms. To this end, two representative networks were seeded with or without  $O_2$  (oxic and anoxic networks, respectively), along with a putatively prebiotic

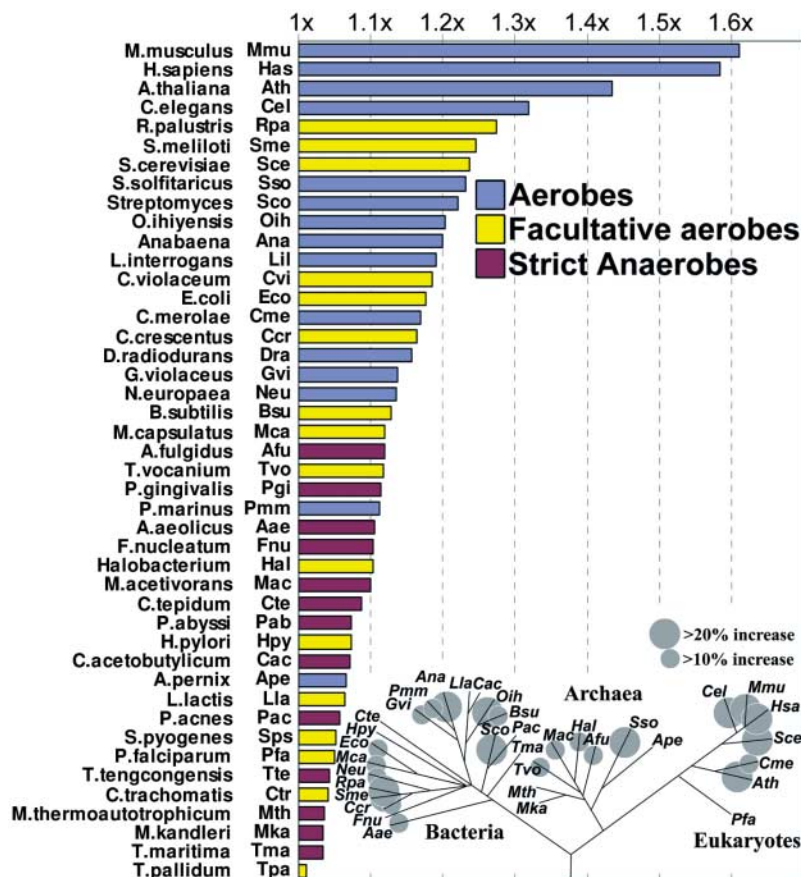


**Fig. 3. (A)** Similarity in anoxic network enzyme distribution in 44 different genomes, fit to two dimensions (see SOM) and broadly consistent with rRNA- and genome-based phylogenies. Blue points are obligate aerobes, yellow points are facultative aerobes that also have anaerobic modes of growth, and maroon points are strict anaerobes. Organism abbreviations are given in Fig. 4. **(B)** Similarity in oxic network enzyme distribution in 44 different genomes, largely inconsistent with organismal phylogeny but, as illustrated, following a trend consistent with an aerobic versus anaerobic growth mode (see also Fig. 4). Axes are in arbitrary units, with closely spaced points representing genomes with greater similarities in enzyme content.

set of metabolites, including  $NH_3$ ,  $H_2S$ ,  $CO_2$ , and the cofactors pyridoxal phosphate, ATP/ADP, tetrahydrofuran (THF), and  $NAD^+/H$ . The expanded anoxic network (Fig. 2, blue links) encompassed 2162 distinct reactions involving 1672 metabolites, analogous to a group III network. The detection of homologs in complete genomes allowed recovery of the distribution of enzymes catalyzing these 1672 reactions in 44 prokaryotic and eukaryotic genomes. Clustering organisms by shared patterns of enzyme distribution (Fig. 3A) reveals an overall pattern consistent with ribosomal RNA (rRNA)- and genome-based phylogenies.

The oxic network resulted in a very different topology and enzyme distribution (Figs. 2 and 3B), spanning 3283 total reactions with 2317 metabolites, which is consistent with a group IV network. A common core of reactions was shared with the anoxic set, and this core was subtracted out to reveal novel and augmented pathways, which occur largely at the periphery of the network (Fig. 2, red, and table S2). A total of 747 reactions and 645 metabolites comprised pathways specific to the oxic network, whereas 273 metabolites were present in both networks but had an augmented set

of reactions with  $O_2$  available (Fig. 2, green). These 273 augmented reactions are an extension of those highlighted by (8) as plausible convergent enzyme replacements: enzymes in anaerobic organisms whose role has been supplanted or replaced by  $O_2$ -dependent enzymes in aerobes. Of the 747 oxic network reactions, only 359 explicitly used molecular oxygen, confirming that metabolic expansion is not simply a result of the invention of  $O_2$ -dependent enzymes but involves the development of novel pathways, such as those shown in table S2 and the multistep pathways in red in Fig. 2. Although the total number of reactions and metabolites increased by 1.5-fold in the oxic network, the density of the network increased only slightly, exemplified by a decrease in path length between any two metabolites from 4.55 to 4.22 in the anoxic and oxic networks, respectively. This suggests that although some pathway “rewiring” indeed occurred after  $O_2$  became available, the most prolific change was in the evolution of new reactions and pathways, making available an entirely new set of metabolites. “Clickable” representations of Fig. 2 and oxic/anoxic reaction databases are available online at <http://prelude.bu.edu/O2/>.



**Fig. 4.** Increase in total number of reactions catalyzed by individual genomes after the inclusion of  $O_2$  in a network originally seeded with  $N_2$ ,  $H_2S$ ,  $CO_2$ , and the cofactors pyridoxal phosphate, ATP/ADP, THF, and  $NAD^+/H$ . Horizontal bars represent the percent increase in oxic versus anoxic network size on a genome-by-genome basis, colored according to the growth mode of the organism (colors are consistent with those in Fig. 3). The inset superimposes this data on a species tree for these organisms, showing that adaptation to  $O_2$  has occurred throughout the tree of life. Also shown are the three-letter KEGG binomen abbreviations used in the inset and in Fig. 3. Enzymes specific to the oxic network expansion observed in strict anaerobes are detailed in SOM and table S3.

By performing a multidimensional scaling analysis of enzyme distribution across different organisms (see fig. S4 for the procedure), we found that the distributions of enzymes catalyzing oxic networks were similar only for closely related organisms, which is inconsistent with the tree of life (Fig. 4, inset) and with the distribution of anoxic enzymes (Fig. 3B). Despite limited data from strict anaerobes, the underlying trend for most organisms is consistent with their preferred aerobic versus anaerobic lifestyle as opposed to taxonomic relationships (Figs. 3B and 4). For example, whereas the three-domain structure of the tree of life is clearly evident in Fig. 3A, aerobes such as the archaeon *Solfolobus solfataricus* and the cyanobacterium *Anabaena* instead cluster with other aerobes, independent of taxonomic domain, in Fig. 3B. Oxic network expansion was most prolific in eukaryotes and aerobic prokaryotes (Fig. 4). Eukaryote-specific reactions make up roughly half of the oxic network—versus 21% of the anoxic network—and among these expanded path-

ways are examples important and in some cases specific to plants and metazoans, such as steroid and alkaloid biosynthesis (table S2).

The fact that enzyme distribution in aerobic pathways was largely incongruent with organismal speciation suggests that adaptation to molecular oxygen occurred after the major prokaryotic divergences on the tree of life. This is supported by data from geological and molecular evolutionary analyses, showing that all three domains of life, and many phyla within these domains, had appeared by the time that oxygen became widely available (20–22). The relatively late onset of atmospheric oxidation argues against the invention of  $O_2$ -dependent enzymes or pathways in the last common ancestor of modern organisms, suggesting that adaptation to molecular oxygen took place independently in organisms from diverse lineages exposed to  $O_2$ . Our data support the idea that  $O_2$  availability is coupled to an increase in network complexity beyond that reachable by any anoxic network, and they highlight enzymes and pathways that might

have been important in the adaptation to an oxic atmosphere and the subsequent development of multicellular life. Whether this enabled the concurrent increase in biological complexity evident in the geological record, or simply was a result of subsequent adaptation and evolution among new classes of aerobes, is of considerable interest but remains to be determined.

#### References and Notes

1. J. L. Kirschvink *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1400 (2000).
2. A. H. Knoll, *Geobiology* **1**, 3 (2003).
3. D. C. Rees, J. B. Howard, *Science* **300**, 929 (2003).
4. J. Raymond, R. E. Blankenship, *Geobiology* **2**, 199 (2004).
5. G. C. Dismukes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2170 (2001).
6. D. J. Des Marais, *Science* **289**, 1703 (2000).
7. R. Rye, H. D. Holland, *Am. J. Sci.* **298**, 621 (1998).
8. R. E. Summons, L. L. Jahnke, J. M. Hope, G. A. Logan, *Nature* **400**, 554 (1999).
9. J. Farquhar, H. Bao, M. Thiemens, *Science* **289**, 756 (2000).
10. O. Ebenhov, T. Handorf, R. Heinrich, *Genome Informatics* **15**, 35 (2004).
11. M. Kanehisa, S. Goto, *Nucleic Acids Res.* **28**, 27 (2000).
12. Briefly, the heuristic can be expressed as (i) specify or randomly choose seed metabolites, (ii) allow metabolites to react according to KEGG reactions, and (iii) incorporate newly synthesized metabolites into seed set. Steps ii and iii are then iterated until convergence, when no further new reactions are possible.
13. During manuscript review and revision, an updated KEGG database (release 37, available 4 January 2006) became available. Results from this latest release are consistent with our findings and are presented along with the SOM.
14. D. Segre, A. Deluna, G. M. Church, R. Kishony, *Nat. Genet.* **37**, 77 (2005).
15. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A. L. Barabasi, *Science* **297**, 1551 (2002).
16. R. Pastor-Satorras, E. Smith, R. V. Sole, *J. Theor. Biol.* **222**, 199 (2003).
17. T. Handorf, O. Ebenhov, R. Heinrich, *J. Mol. Evol.* **61**, 498 (2005).
18. H. Morowitz, *Complexity* **4**, 39 (1999).
19. O. Ebenhov, T. Handorf, R. Heinrich, *Genome Inform. Ser. Workshop Genome Inform.* **16**, 203 (2005).
20. J. Xiong, W. M. Fischer, K. Inoue, M. Nakahara, C. E. Bauer, *Science* **289**, 1724 (2000).
21. J. J. Brocks, G. A. Logan, R. Buick, R. E. Summons, *Science* **285**, 1033 (1999).
22. J. M. Hayes, *Global Methanotrophy at the Archean-Proterozoic Transition, Early Life on Earth, Nobel Symposium No. 84*, S. Bengtson, Ed. (Columbia Univ. Press, New York, 1994).
23. We thank R. Blankenship and E. Branscomb for critical review and fruitful discussions. J.R. acknowledges support through the E. O. Lawrence Fellowship at Lawrence Livermore National Laboratory (LLNL). D.S. is supported at LLNL as a Faculty Scholar. This work was performed under the auspices of the U.S. Department of Energy at LLNL, contract no. W-7405-ENG-48.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/311/5768/1764/DC1  
SOM Text  
Figs. S1 to S4  
Tables S1 to S3  
References

4 August 2005; accepted 1 February 2006  
10.1126/science.1118439