Molecular Phylogenetics and Evolution 66 (2013) 442–452

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

The lemur revolution starts now: The genomic coming of age for a non-model organism

Anne D. Yoder

Department of Biology, Duke University, Duke Lemur Center, Durham, NC 27708, United States

ARTICLE INFO

Article history:
Available online 7 September 2012

Keywords:
Madagascar
Phylogenomics
Biogeography
Divergence time
Primates

ABSTRACT

Morris Goodman was a revolutionary. Together with a mere handful of like-minded scientists, Morris established himself as a leader in the molecular phylogenetic revolution of the 1960s. The effects of this revolution are most evident in this journal, which he founded in 1992. Happily for lemur biologists, one of Morris Goodman's primary interests was in reconstructing the phylogeny of the primates, including the tooth-combed Lorisifomes of Africa and Asia, and the Lemuriformes of Madagascar (collectively referred to as the suborder Strepsirrhini). This paper traces the development of molecular phylogenetic and evolutionary genetic trends and methods over the 50-year expanse of Morris Goodman's career, particularly as they apply to our understanding of lemuriform phylogeny, biogeography, and biology. Notably, this perspective reveals that the lemuriform genome is sufficiently rich in phylogenetic signal such that the very earliest molecular phylogenetic studies – many of which were conducted by Goodman himself – have been validated by contemporary studies that have exploited advanced computational methods applied to phylogenomic scale data; studies that were beyond imagining in the earliest days of phylogeny reconstruction. Nonetheless, the frontier still beckons. New technologies for gathering and analyzing genomic data will allow investigators to build upon what can now be considered a nearly-known phylogeny of the Lemuriformes in order to ask innovative questions about the evolutionary mechanisms that generate and maintain the extraordinary breadth and depth of biological diversity within this remarkable clade of primates.

> Where you work and where you play
> Where you lay your money down
> What you do and what you say
> The revolution starts now
> **STEVE EARLE – THE REVOLUTION STARTS NOW LYRICS**

1. Introduction

Phylogeneticists of all stripes glory in the fact that Darwin chose to illustrate a phylogeny as the sole figure in *The Origin of Species*. It is also well established that Darwin ended his days without ever knowing what is the biological mechanism of heritable variability (though see a very thoughtful essay by Charlesworth and Charlesworth, 2009 on what Darwin did and did not surmise about heritability). Copious books, essays, and empirical accounts have been written on the first meeting of Mendelian genetics and macroevolutionary thought, yielding the great evolutionary synthesis of the 1930s and 1940s. It was then, finally, that Darwinian perspectives on phylogeny began to take an indelible hold on biological thought. From that moment onwards, it has been the unrelenting goal of phylogeneticists to assemble this grand Tree of Life.

Beginning in the 1950s with protein electrophoresis, molecular biologists started to tinker with the idea that measures of genetic distance among and between organisms could be interpreted as a proxy for their evolutionary relatedness. The obvious thought was that organisms that share the most recent ancestry will show the greatest similarity of genetic material. Pioneering work by Walter Fitch, Emile Zuckerkandl, and Linus Pauling laid the groundwork for the molecular phylogenetic revolution, led principally by Allan Wilson, and by Morris Goodman, to whom this special volume is dedicated. Although the first decade or so of this revolution relied upon indirect measures of genetic distance such as DNA–DNA hybridization, numerous breakthroughs in our understanding of evolutionary relationships were achieved, such as the (very controversial, at the time) finding that chimpanzees are more closely related to humans than to gorillas (Sibley and Ahlquist, 1984). Several authors took exception to these results in particular, both in terms of the obvious incongruence with the morphological details shared by chimpanzees and gorillas (reviewed in Holmquist et al., 1988), but also due to various subtleties of statistical analysis (Farris, 1985; Templeton, 1985).

E-mail address: anne.yoder@duke.edu

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.08.024

The dispute was for many settled decisively by Felsenstein (1987) who employed a maximum likelihood mixed model analysis of variance method to show that there was indeed significant support for the human-chimp clade contained within the DNA–DNA hybridization data published by Sibley and Ahlquist, 1984. Felsenstein pursued the matter further by exploring the question of just how many base pairs of DNA sequence data would convey the same degree of statistical power as the vast amount of genetic material being compared by hybridizing the single-copy regions of whole genomes. His answer was very precise: 4472 base pairs of DNA sequence data would convey equivalent power. This result would have come as no surprise to Morris Goodman and the other molecular phylogenetic revolutionaries who had long been utilizing amino acid sequence data for resolving questions of evolutionary relatedness (Goodman et al., 1972, 1974; Matsuda et al., 1973; Moore et al., 1973).

The molecular phylogenetics field moved rapidly and nearly uniformly to the analysis of DNA sequence data coincident with the PCR revolution launched by Kary Mullis (Mullis et al., 1986), and for many years, the field has been driven nearly exclusively by PCR and Sanger sequencing based methods. Studies have evolved from sampling strategies in which only a few taxa were sequenced for only one organellar or nuclear locus, to combined analysis of representative loci from both genomes, to whole mitochondria (i.e., mitogenomics) to large-scale samples of nuclear loci (i.e., phylogenomics). Founded by Morris Goodman in 1992, the journal *Molecular Phylogenetics and Evolution* was created specifically to “disseminate the results of these molecular studies” (Goodman, 1992). This dream has been more than fulfilled, and indeed, a backwards glance at the content of the journal can be viewed as a mirror of the developing molecular phylogenetic field (Table 1) which is today undergoing its latest and perhaps greatest revolution. Although in the year of its founding, the field was comprised by phylogenetic information that was “miniscule compared to the huge reservoirs that remain[ed] to be tapped” (Goodman, 1992), the progress of the past few years is truly astounding. Table 1, which is a tabulation of basic information from each year of the journal, clearly illustrates this progress. Whereas studies of 20 years ago tended to rely on parsimony or distance-based analysis of small subsets of genetic data and OTUs (operational taxonomic units), there has been a steadily increasing trend towards more loci, more OTUs, and increasingly sophisticated statistical analysis of the data (Table 2). Most notably, the journal has had to increase the number of published papers by more than a factor of ten to keep up with the outpouring of empirical and methodological studies. Clearly, each technological advance in data generation has been quickly followed by studies with increasing amounts of data, which in turn have necessitated analytical methods and tools of increasing statistical and computational power. We see a version of the Red Queen Hypothesis played out in the pages of *Molecular Phylogenetics* and *Evolution*.

With the advent of “next generation” sequencing methods first introduced in 2005 (see Egan et al., 2012 for a detailed history of these technologies), it is now possible to generate millions of bases at a fraction of the cost of traditional Sanger methods. Accordingly, the field is starting to move rapidly in the direction of whole genome sequencing, not only for the purposes of resolving evolutionary relationships, but for any conceivable application of genomic data to fields as disparate as molecular ecology and cancer biology. Happily, Morris Goodman not only lived to see these advances, he was fully immersed in their applications (Goodman and Sterner, 2010; Goodman et al., 2009, 2010; Jameson et al., 2011; Sterner et al., 2010).

2. Phylogeny of the lemurs: Nearly known

Lemurs have been the focus of molecular phylogenetic study from the earliest days of the field’s emergence. The suborder Lemuriformes is comprised entirely of primate species endemic

Table 1
MPE publication trends.

Year	Data				# of OTUs			Phylogenetic analysis				Journal statistics		
	organell only (mtDNA or cpDNA)	nDNA only	Organelle & nDNA	Whole mtDNA genomes	Minimum #	Maximum#	Mean #	Parsimony	Distance	Likelihood	Bayesian	# of volumes	# of issues	# of papers
1992	7	3	0	0	5	14	10	7	3	3	0	1	4	31
1993	2	8	0	0	4	47	23	5	8	1	0	1	4	35
1994	6	3	0	0	5	36	15	6	4	0	0	1	4	40
1995	5	3	2	0	7	56	21	8	8	3	0	1	4	43
1996	3	6	0	1	8	34	20	8	5	2	0	2	6	84
1997	6	3	1	0	8	42	25	9	8	1	0	2	6	72
1998	4	5	1	0	10	60	28	9	8	3	0	2	6	101
1999	6	4	0	0	7	49	28	9	8	6	0	3	9	123
2000	4	5	1	0	14	49	31	9	5	6	0	4	12	167
2001	7	0	3	0	14	67	37	10	6	10	0	4	12	166
2002	7	2	1	0	12	78	39	10	3	4	0	4	12	162
2003	8	1	1	0	13	165	45	10	1	7	1	4	12	189
2004	6	4	0	0	20	100	46	9	1	6	6	4	12	332
2005	4	2	3	1	21	142	66	8	3	8	8	4	12	234
2006	5	1	3	1	13	834	131	8	2	6	8	4	12	276
2007	3	1	5	1	15	76	42	9	1	5	5	4	12	345
2008	3	3	3	1	23	136	70	9	3	4	8	4	12	387
2009	0	2	7	1	22	161	78	7	1	4	8	4	12	297
2010	1	3	6	0	42	102	69	5	1	3	9	4	12	431
2011	0	1	8	1	18	282	88	6	1	7	8	4	12	248
2012	1	3	6	0	8	241	98	5	2	8	8	TBD	TBD	TBD

Represents sample of first 10 empirical studies (regardless of organismal focus) from each journal year. Data type, # of OTUs and phylogenetic methods were tabulated from empirical studies only (i.e., simulation or method development studies were not considered); symposium proceedings were also not considered in tabulation of these statistics due to potential bias; the majority of empirical studies used multiple optimality criteria, and thus do not sum to 10; # of papers does not include editorial remarks, book reviews or eratta. The author does not claim precise accuracy of the data, but stands by the observable trends described in the body of the paper.