Journal of Phylogenetics & Evolutionary Biology

Felix B, J Phylogen Evolution Biol 2015, 3:2
DOI: 10.4172/2329-9002.1000e112

Editorial Open Access

Phylogenetics: Tracing the Evolutionary Legacy of Organisms, Metastatic Clones, Bioactive Compounds and Languages

Felix B*

Centre for Biosciences, Central University of Punjab, Mansa Road, Bathinda, 151001, Punjab, India

Abstract

Since its inception seventy-five years ago, the field of phylogenetics has steadily been expanding to contribute in a number of scientific fields including biogeography, medicinal chemistry, forensics, transcriptomics, cancer biology and even linguistics, in addition to systematic biology -for which it was originally erupted by Willi Hennig. In this invited editorial contributed to the Journal of Phylogenetics and Evolutionary Biology a big picture of this ever evolving field is expounded. Application of phylogenetic inference in biosystematics, phylogeography, phylogenetic selection of target taxa in medicinal chemistry, cancer phylogenetics, and linguistic phylogeny are reviewed with a personal perspective summarizing contribution to this interdisciplinary field from my group. Parametric methods such as Maximum Likelihood and Bayesian Inference have dramatically improved for last one decade, yet empirical solutions to some of the most fundamental issues, including homoplasmy and lineage sorting, remains to be materialized.

Keywords: Cancer phylogenetics; Coalescence; Gene tree; Historical linguistics; Medicinal chemistry; Phylogeny; Phylogeography; Tree of life

"Time is a sort of river of passing events, and strong is its current"

-Marcus Aurelius -16th Roman Emperor, Stoic Sage and Philosopher

This oft-quoted statement, though primarily intended to enlighten our consciousness to live in this fleeting presence, indeed reverberate over three billion years old odyssey of the life on planet earth. Tracing those passing events to make sense of Darwin's "tree of life" is what comes in the realms of phylogenetics-an arena that combined the power of probabilistic statistics with evolutionary biology. As simple as it may seem, phylogenetic inference starts with one simple but profound premise: past informs the present.

Originally conceived in 1950 by German entomologist Willi Hennig [1] who used this technique in systematic taxonomy, the scope of phylogenetics have ever since expanded tremendously to a number of applied scientific and even liberal arts disciplines including cancer biology, medicinal chemistry, linguistics, and forensics. For example, by phylogenetic reconstruction of HIV strains -that evolve with each new infection leaving certain relics of the past, a dentist in Florida was found guilty of deliberately infecting the virus to his patients [2]. Research from my own group have revealed in 2015 that sporadic episodes of the "blood rain" phenomenon, reported since time immemorial and even can be found in Homer's *Iliad*, was due to the spores of subareal green microalgae *Trentepohlia annulata*, and that the strain of this algae from South India and that from Central Europe had such a DNA sequence homology to have it introduced very recently [3]. Currently phylogenetics occupy the center stage of practical evolutionary biology for tracing the evolutionary legacy of varied subjects including the originally-intended biological species, and products of scientific ingenuity such as gene expression profiles of microarray data (transcriptomics), metastatic clones in cancer, bioactive compounds in medicinal chemistry and even languages in comparative and historical linguistics.

The field of systematic taxonomy have greatly benefitted from phylogenetics such that a new field "phylogenetic systematics" had been conceived concurrently with the publication of eponymous work by Hennig in 1965 [4]. Linnaean systematics classified organisms based on overall similarity, an approach known as phenetics. As opposed, Hennigian phylogenetic systematics incorporated Darwinian Theory of evolution in classifying organisms into the tree of life, an approach sometimes referred as cladistics. This field has since then actively contributing in both morphometry-based traditional systematics and modern molecular systematics. Since the advent and widespread use of computers, the field of computational phylogenetics started gaining momentum and arrived at the center-stage of applied phylogenetics. While earlier computational phylogenetics software packages had severe computational constraints that limited their use to the most fundamental of distance matrix-based statistical techniques such as Neighbor-Joining and UPGMA, thanks to the ever-increasing computational speed quite in accordance with the Moore's law that today's packages are able to implement discrete character-based techniques, that include non-parametric statistics (parsimony) and parametric statistics (such as maximum likelihood and Bayesian Inference). A number of excellent computational phylogenetic packages are now available across operating systems such as Geneious, PAUP, MEGA and BEAST that implement both parametric and non-parametric phylogenetic frameworks. Many of these packages are exhaustive such that a user can use it right from base-calling of their DNA electropherograms, contig sequence assembly, homology search, multiple sequence alignment, concatenation and consensus sequence generation to model selection and phylogenetic inference [5]. With more and more whole genome sequences are made available -thanks to the fast evolving NextGen sequencing technologies, the field of phylogenomics started to emerge. However, given that many phylogeneticists are already working with whole genome data, need for pruning out such a field might be nothing but mere fancy. Using phylogenetic systematics approach our group were able to report existence of sympatric speciation in panmictic population of eukaryote (in *Monostroma kuroshiense* Bast a marine green algae found in SW Japan) for the first time [6]. We also discovered two species of bloom-forming marine green algae endemic to India; *Ulva paschima* Bast and *Cladophora goensis* Bast [7,8], and reported for the first time the occurrence of endophytic algae in Indian Ocean [9].

A related field conceived by none other than the contemporary of Darwin and co-discoverer of natural selection, Alfred Russel Wallace, is biogeography that deals with distribution of species and ecosystems in space through deep geological time. Phylogenetic methods are now

***Corresponding author:** Felix B, Centre for Biosciences, Central University of Punjab, Mansa Road, Bathinda, 151001, Punjab, India, E-mail: felix.bast@gmail.com

Received March 23, 2015; **Accepted** March 25, 2015; **Published** March 30, 2015

Citation: Felix B (2015) Phylogenetics: Tracing the Evolutionary Legacy of Organisms, Metastatic Clones, Bioactive Compounds and Languages. J Phylogen Evolution Biol 3: e112. doi:10.4172/2329-9002.1000e112

routinely used to trace patterns and routes of species' dispersal, as well as to identify geographic origin of a species, that lead to synthetic new discipline of phylogeography[10]. Modern phylogeography often combines other biogeographical evidences such as satellite data retrieved using GIS (Geographic Information Systems) to generate a concatenated dataset featuring "total evidence", not merely restricting to DNA sequence data. Research in this direction enabled our own group to trace geographical origin of cultivated edible green algae in Japan, *Monostroma kuroshiense* Bast, to Ise Bay in Central Japan [6] and holy basil in India, *Ocimum tenuiflorum* L to North-Central India [11].

Biogeography is surprisingly analogous to clonal evolution in metastatic cancers; like rapidly multiplying lineages of adaptive radiations across islands in an archipelago, clones of tumor biomass is evolving across multiple neoplastic progression [12]. Many of our findings in cancer phylogenetics credit to the fact that the tumor is not merely a collection of transformed cells with random mutation events; rather it is an evolving population. Many of the facets underpinning modern evolutionary synthesis can be applied to classify cancers and track its progression from initiating somatic mutation to symptomatic neoplasm. It is now widely accepted that all sub-clones within cancer are phylogenetically related and probability of a particular sub-clone progressing into neoplasm depending upon its time of initiation and evolutionary fitness. Computational models of tumor evolution have also contributed in identifying common clades- "cancer sub-types"- associated with particular cancers in different patients that in turn helped in translating our understanding of oncogeny to the development of "targeted therapeutics"- rationally designed drugs that are molecularly targeted to particular sub-types. Based on the presence or absence of specific mutation events, microarray gene expression measurements and gene copy numbers measured by Comparative Genomic Hybridization (CGH), it was Desper et al. who applied the principles of phylogenetics for the first time in the field of cancer to calculate evolutionary distances for inferring inter-tumor *oncogenetic trees* [13]. Multiple genetically related subclones in a case of B-Cell Chronic Lymphocytic Leukemia and clonal interrelationships within were demonstrated by phylogenetic analyzes aided by massively parallel *de novo* DNA sequencing methods [14]. Tumors preserve remnants of earlier cell populations as they develop [15,16]. Tumors, therefore, contain transformed cells at various stages of progression as well as healthy "contaminating" cells [12].

Another field that has surprisingly benefitted from phylogenetics is medicinal chemistry. Until recently screening of natural products for bioactivity had been a herculean task, sometimes needed to screen tens of thousands of extracts from random plant species. This "trial and error" approach is indeed very laborious and expensive. For example, a large number of (16,000) marine species were screened for anticancer potentials at the National cancer Institute, USA over 22 year period (1960-1982), but the program was subsequently abandoned as only a few chemical leads were discovered. Recently, a new approach- "phylogenetic selection of target taxa"- has been emerging inspired from the evolutionary theory. A number of investigations have revealed that bioactive substances (secondary metabolites) are synthesized by organisms as an adaptive strategy for differential reproductive success and ultimately for long-term evolution of the fittest populations. Evolution of bioactive substances is, therefore, most likely to be in parallel with the evolution of species themselves. In this context, the evolutionary legacy, as traced by computational phylogenetics, could be effectively used as a guide for choosing the target taxa, thereby making the target selection way more natural and smarter, as opposed to trial and error approach that is traditionally being employed. A correlation between phylogeny and biosynthetic pathways could offer a predictive approach enabling more efficient selection of plants for the development of traditional medicine and lead discovery.

Phylogenetic selection of target species has been successfully tested in Amaryllidaceae-a family of herbaceous, perennial and bulbous flowering plants in the monocot order Asparagales. In one study, the correlation between phylogenetic and chemical diversity and biological activity had been concluded [17]. The study concluded that phylogenies have potential to interpret chemical evolution and biosynthetic pathways, to select candidate taxa for lead discovery, and to make recommendations for policies regarding traditional use and conservation priorities. In another study in the same family, a clade (Clade-D) that had the longest branch length, had been identified as a lineage that might possess derived alkaloid chemistry and potentially new and more active biomolecules [18]. The same study has also identified few phylogenetic clades that corresponded very low IC-50 values and concluded that these clades were "uninteresting" as a candidate for future attempts for the discovery of novel bioactive agents. Phylogenetic selection have also successfully attempted to discover acetylcholinesterase inhibiting alkaloids in Amaryllidaceae[18] and Narcissus –a Mediterranean perennial geophyte [19]. A recent meta-analysis of marine metazoans collected from Australia concluded that the phylogenetic relatedness is the primary determinant of the level of bioactivity due to the evolution of similar metabolic pathways [20]. The study identified deuterostome's lineage as the most promising metazoan lineage for the discovery of bioactive agents. Phylogenetics have also been successfully used to determine the target clades of cyanobacteria that produce natural products [21]. A study that constructed genus-level phylogenetic tree representing 20,000 species from three biodiversity hotspots revealed that traditional medicinal plant use is not scattered randomly but is concentrated in certain parts of the phylogenetic trees [22]. This study demonstrates the power of phylogeny in predicting medicinally significant taxonomic groups. Our group is currently involved with phylogenetic selection of target marine algal taxa for anti-metastatic marine natural compounds-a research supported by Drugs from the Sea program of Ministry of Earth Sciences, Government of India.

Many of the fundamental features and concepts of biological and linguistic evolution are remarkably analogous that made Charles Darwin conceptualize these (biological vs. linguistic evolution) as "Curious Parallels." In *The Descent of Man*, Darwin (1871) noted that the process of evolution is not confined to just the biological realm.

"The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel… we find in distinct languages striking homologies due to community of descent, and analogies due to a similar process of formation."

Just like biological systems, languages too are descended from generation to generation; as homologous structures in animals indicating inheritance from a common ancestor, languages too have homologous words (cognates). Evolutionary processes like mutation (innovation) and horizontal gene transfer (borrowing) have parallel phenomena in linguistics, as well. Phenomena like natural selection (social selection) and random drift (linguistic drift) also act upon languages, and the formation of new languages and its disappearance remarkably resembles processes of speciation and extinction. This parallelism would mean that statistical methods inspired by phylogenetics and comparative biology are being increasingly applied to study the language. While inferring and representing evolutionary heritage of life using statistical tools is the primary focus of phylogenetics, the field has immense potentials in historical linguistics owing to the

curious parallels. Linguistic phylogenetics can also compliment human evolutionary phylogenetics by providing crucial information on the migration (including immigration and emigration) patterns.

In a pioneering paper published in 1988, LL Cavalli-Sforza et al. [23] presented an illustration directly comparing a human linguistic and genetic tree. Albeit extremely controversial, the Cavalli-Sforza et al. paper has highlighted the similarities between processes of historical inference in biology and linguistics, as well as the potential importance of linguistic data for inferences about human population history. In the wake of this paper, there has been a surge of studies attempting to test hypotheses about human population history and something of a resurgence of interest in computational phylogenetic methods, in historical linguistics. This "new synthesis" of biology and linguistics has provided solutions to many of the problems that plagued lexicostatistics and glottochronology. For example, character-based tree- building techniques retain individual character state information, thus avoiding the problem of information loss associated with distance-based methods. In a phylogenetic analysis, the evolutionary legacy is traced to a set of subjects; subjects are biological species in biology and languages, or dialects, in linguistics. A set of characters common to all subjects are taken as a dataset in both of these analyzes, and each subject is represented by the states for these characters. Discrete characters for linguistic phylogenetics could include the lexicon, syntax and phonology –using them one can create data sets of languages that can be analyzed in the robust framework of computational phylogenetics to trace the linguistic evolutionary legacy.

Phylogenetic methodologies have been routinely employed for the linguistic analyzes of a number of languages. In one of the well-received studies, linguistic phylogeny of Oceanic Austronesian languages were conducted using cladistic analysis of a dataset consisting of structural elements (sound systems and grammar) revealed ancient lineage split of Papuan languages from the main group [24]. Bayesian analyzes of Semitic languages revealed this language family had been originated in early bronze-age in Near East [25]. Indo-European alone [26] or along with Gaulish and Celtic [27] linguistic datasets were also subjected to the comparative phylogenetics, albeit objectives of these studies were to compare phylogenetic methods. D Ringe, T Warnow and A Taylor [28] used compatibility methods to infer an Indo-European language tree from discrete grammatical, lexical and phonological characters. RD Gray and FM Jordan [29] conducted a parsimony analysis of over 5000 discrete lexical characters to find an optimal tree for 77 Austronesian languages. They then used this tree to test competing scenarios for the settlement of the Pacific. CJ Holden [30] applied similar methods to test migration scenarios in the Bantu language family in Africa and K Rexová, D Frynta and J Zrzavý [31] constructed an Indo-European language tree, also using parsimony methods. Our group is currently involved in tracing the linguistic evolutionary legacy of Indian languages using computational linguistics-a research supported by Indian Council for Social Science Research.

Conclusions

The field of phylogenetics does have a number of challenges; for example ever increasing debate between parametric and non-parametric statisticians over which approach is more robust. Or whether a coalescent-based retrospective approach that are being favored by population geneticists- are even more superior to phylogenetic based methods. Incongruence between topologies of gene trees and species trees- the problem of lineage sorting- still haunt almost every single phylogenetic tree published. Despite all these short comings it is expected that the field of phylogenetics will continue to light-up scientific fields that are enlightened by the theory of evolution. To put a step forward to Dobzansky's posit, nothing in Life and Language make sense except in the light of evolution.

Acknowledgements

This work is supported in part by grants in aid from INSPIRE Faculty Award IFA11/LSPA02 from Department of Science and Technology, Government of India and Major Research Project GP-35 from Indian Council of Social Science Research, Government of India.

References

1. Hennig W (1950) Grundzugeeiner Theorie der phylogenetischen Systematik (1stedn), Berlin: Deutscher Zentralverlag, Arlington, MA, U S A pp: 1
2. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, et al. (1992) Molecular epidemiology of HIV transmission in a dental practice. Science 256: 1165-1171.
3. Bast F, Bhushan S, John AA, Achankunju J, Panikkar NMV, et al. (2015) European Species of Subaerial Green Alga *Trentepohlia annulata* (Trentepohliales, Ulvophyceae) Caused Blood Rain in Kerala, India. J Phylogen Evolution Biol3:144.
4. Hennig W (1965) Phylogenetic systematics. Annual review of entomology 1:97-116.
5. Bast F (2013) Sequence Similarity Search, Multiple Sequence Alignment, Model Selection, Distance Matrix and Phylogeny Reconstruction. Nature Protocol Exchange.
6. Bast F, Kubota S, Okuda K (2015) Phylogeographic Assessment of PanmicticMonostroma Species from Kuroshio Coast, Japan Reveals Sympatric Speciation. J Appl Phycol.
7. Bast F, John AA, Bhushan S (2014) Strong Endemism of bloom-forming tubular Ulva in Indian West Coast, with description of Ulvapaschima Sp. Nov. (Ulvales, Chlorophyta). PLoS One 9: e109295.
8. Bast F, John AA, Bhushan S (2014) Cladophoragoensis Sp. Nov. (Cladophorales, Ulvophyceae) –a bloom forming marine algae from Goa, India. Indian J of Geo-Marine Sciences.
9. Bast F, Bhushan S, John AA (2014) DNA barcoding of a new record of epi-endophytic green algae Ulvellaleptochaete (Ulvellaceae, Chlorophyta) in India. J Biosci 39: 711-716.
10. Huang Z, Liu N, Luo S, Long J (2007) Phylogeography of rusty-necklaced partridge (Alectoris magna) in northwestern China. Mol Phylogenet Evol 43: 379-385.
11. Bast F, Rani P, Meena D (2014) Chloroplast DNA phylogeography of holy basil (*Ocimum tenuiflorum*) in Indian subcontinent. Scientific World Journal 2014: 847482.
12. Bast F (2012) Cancer Phylogenetics: Computational Modeling of Tumor Evolution. In: Bioinformatics: Genome Bioinformatics and Computational Biology, Nova Publishers, New York, pp: 211-230.
13. Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, et al. (2000) Distance-based reconstruction of tree models for oncogenesis. J Comput Biol 7: 789-803.
14. Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, et al. (2008) Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. ProcNatlAcadSci U S A 105: 13081-13086.
15. Pennington G, Smith CA, Shackney S, Schwartz R (2007) Reconstructing tumor phylogenies from heterogeneous single-cell data. J Bioinform Comput Biol 5: 407-427.
16. Pennington G, Smith CA, Shackney S, Schwartz R (2006) Expectation-maximization method for reconstructing tumor phylogenies from single-cell data. ComputSyst Bioinformatics Conf .
17. Rønsted N, Symonds MR, Birkholm T, Christensen SB, Meerow AW, et al. (2012) Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of amaryllidaceae. BMC evolutionary biology 12: 182.
18. Larsen MM, Adsersen A, Davis AP, Lledó MD, Jäger AK, et al. (2010) Using a phylogenetic approach to selection of target plants in drug discovery of acetylcholinesterase inhibiting alkaloids in Amaryllidaceae tribe Galantheae. Biochemical systematics and ecology 38:1026-1034.
19. Rønsted N, Savolainen V, Mølgaard P, Jäger AK (2008) Phylogenetic selection of Narcissus species for drug discovery. Biochemical systematics and ecology 36:417-422.

20. Evans-Illidge EA1, Logan M, Doyle J, Fromont J, Battershill CN, et al. (2013) Phylogeny drives large scale patterns in Australian marine bioactivity and provides a new chemical ecology rationale for future biodiscovery. PLoS One 8: e73800.

21. Engene N1, Gunasekera SP, Gerwick WH, Paul VJ (2013) Phylogenetic inferences reveal a large extent of novel biodiversity in chemically rich tropical marine cyanobacteria. Appl Environ Microbiol 79: 1882-1888.

22. Saslis-Lagoudakis CH, Savolainen V, Williamson EM, Forest F, Wagstaff SJ, et al. (2012) Phylogenies reveal predictive power of traditional medicine in bioprospecting. Proceedings of the National Academy of Sciences 109: 15835-15840.

23. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci U S A 85: 6002-6006.

24. Dunn M, Terrill A, Reesink G, Foley RA, Levinson SC (2005) Structural phylogenetics and the reconstruction of ancient language history. Science 309: 2072-2075.

25. Kitchen A, Ehret C, Assefa S, Mulligan CJ (2009) Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. Proc Biol Sci 276: 2703-2710.

26. Nakhleh L, Warnow T, Ringe D, Evans SN (2005) A comparison of phylogenetic reconstruction methods on an Indo?European dataset. Transactions of the Philological Society 103: 171-192.

27. Forster P, Toth A (2003) Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European. Proc Natl Acad Sci U S A 100: 9079-9084.

28. Ringe D, Warnow T, Taylor A (2002) Indo-European and Computational Cladistics. Transactions of the Philological Society 100:59-129.

29. Gray RD, Jordan FM (2000) Language trees support the express-train sequence of Austronesian expansion. Nature 405: 1052-1055.

30. Holden CJ (2002) Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. Proceedings of the Royal Society of London Series B: Biological Sciences 269:793-799.

31. Rexová K, Frynta D, Zrzavý J (2003) Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. Cladistics 19:120-127.