Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

The evolution of animal genomes

Casey W Dunn[1] and Joseph F Ryan[2,3]

Genome sequences are now available for hundreds of species sampled across the animal phylogeny, bringing key features of animal genome evolution into sharper focus. The field of animal evolutionary genomics has focused on identifying and classifying the diversity genomic features, reconstructing the history of evolutionary changes in animal genomes, and testing hypotheses about the evolutionary relationships of animals. The grand challenges moving forward are to connect evolutionary changes in genomes with particular evolutionary changes in phenotypes, and to determine which changes are driven by selection. This will require far greater genome sampling both across and within species, extensive phenotype data, a well resolved animal phylogeny, and advances in comparative methods.

Addresses
[1] Department of Ecology and Evolutionary Biology, Brown University, 80 Waterman St., Providence, RI 02906, USA
[2] Whitney Laboratory for Marine Bioscience, University of Florida, 9505 Ocean Shore Blvd., St Augustine, FL 32080, USA
[3] Department of Biology, University of Florida, Gainesville, FL 32611, USA

Corresponding author: Dunn, Casey W (casey_dunn@brown.edu)

Current Opinion in Genetics & Development 2015, **35**:25–32

This review comes from a themed issue on **Genomes and evolution**

Edited by **Antonis Rokas** and **Pamela S Soltis**

http://dx.doi.org/10.1016/j.gde.2015.08.006

The goals of animal evolutionary genomics

Four main goals drive much of the work on animal evolutionary genomics: firstly, to reconstruct the history of animal genome evolution, secondly, to identify evolutionary changes in genomes which contributed to evolutionary changes in phenotypes, thirdly, to understand the evolutionary processes (including selection and neutral drift) that led to genome change, and finally to use the evolution of genomes as a proxy for understanding other historical patterns such as phylogeny and demography. This review examines progress towards these goals, provide examples of biological insights that have already been achieved, and discuss priorities for future work. It focuses on macroevolution of nuclear genomes.

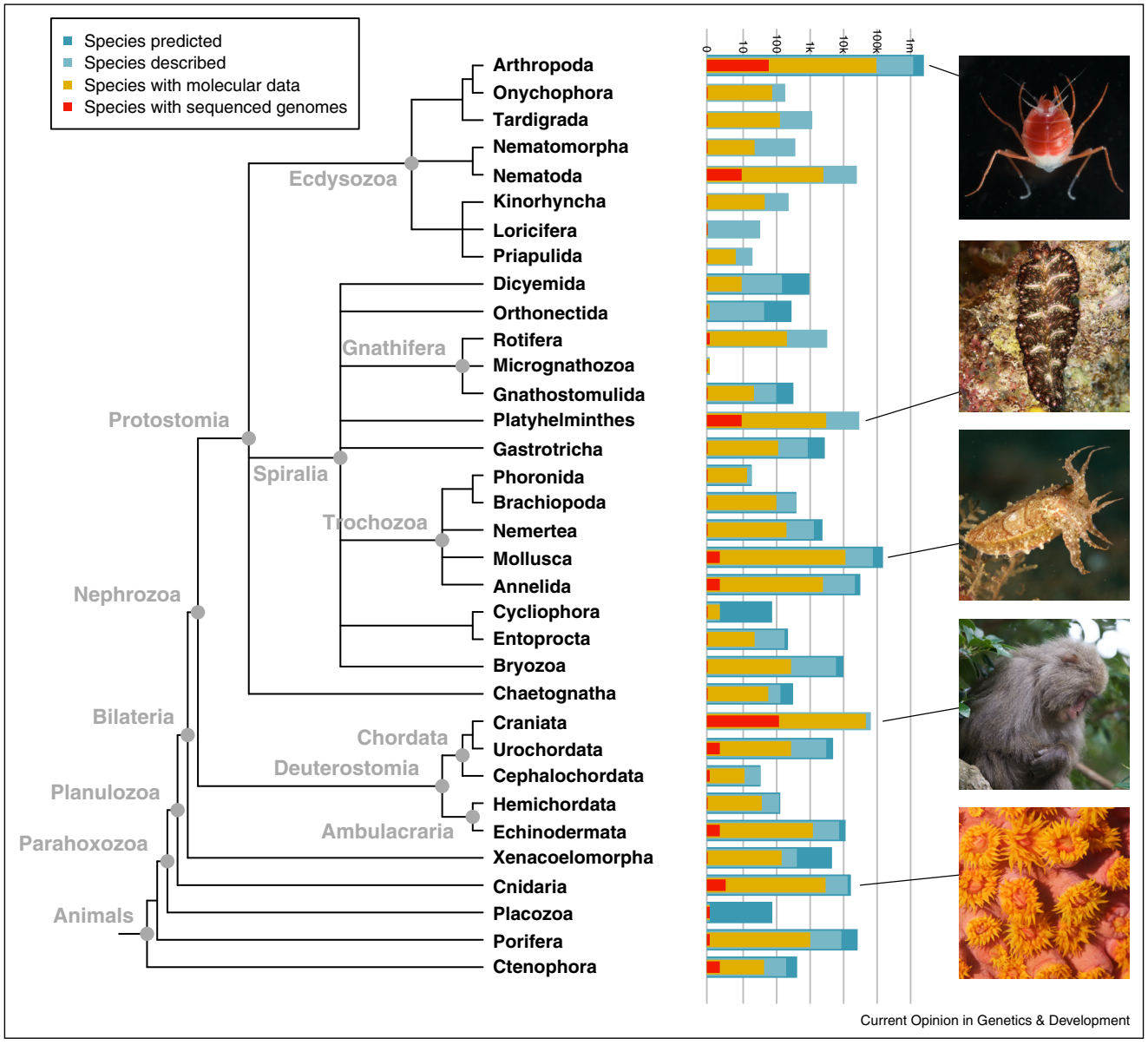
We are still at a very early stage of the genomic revolution, not unlike the state of computing in the late nineties [1]. Genome sequences of various levels of quality are now available for a few hundred animal species thanks to the concerted efforts of large collaborative projects [2–4], more recent efforts by smaller groups [5–7], and even genome projects undertaken largely by independent laboratories [8,9]. At least 212 animal genomes have been published (Figure 1) and many others are publicly available in various stages of assembly and annotation. This sampling is still very biased towards animals with small genomes that can be bred in the laboratory and those from a small set of phylogenetic clades. 83% of published genomes belong to vertebrates and arthropods (Figure 1). Even so, the breadth of sampling is much better than it was a few years ago, with genomes now available for distantly related species that span the root of the animal tree.

All the goals outlined above are advanced by improved taxonomic sampling [10]. Despite progress sequencing a diversity of animal genomes, there is still a long way to go. Genome sequences have been published for less than 0.015% of the roughly 1.5 million described animal species (Figure 1). Many published genome sequences are of poor quality due to the challenges of repetitive DNA, heterozygosity, and contamination [11–13], and some important clades of animals remain entirely unsampled. Improving quality and sampling is a critical and tractable goal, and there are multiple efforts underway to achieve this (e.g., [14•]). As sequencing and analysis methods improve, the limiting steps for gaining biological insight from animal genomes will shift from sequencing expenses and general analysis expertise to the taxon-specific expertise required to collect specimens and interpret results, and the costs associated with obtaining wild animals across the globe. At present, there are 167,802 animal species with molecular sequence data in NCBI (Figure 1), about 11.5% of described species (and about 6% of the animal species thought to live on Earth). These are the species that we know can be collected, identified, and brought back to the lab at current funding levels. To an optimistic first approximation, the taxon sampling of genomes in the next couple decades will be similar to the current sampling of animals for which at least one gene sequence is available.

Reconstructing the history of animal genome evolution

As animal genome sampling has improved, the results have surprised many investigators. It was postulated that animals differ greatly in complexity, and it was expected that gene number would be correlated with these differences [15]. Instead, genome after genome has revealed

Figure 1



The estimated number of species, number of described species [14•], number of species with any publicly available molecular sequence data (as listed in NCBI on May 14, 2015), and number of species with sequenced nuclear genomes (parsed from http://wikipedia.org/wiki/List_of_sequenced_animal_genomes on June 4, 2015, which we updated) across the animal phylogeny. The topology of the phylogeny is from [31]. The estimated numbers of species for marine clades were taken by averaging the ranges in [63], for Arthropoda from [64]. The source code and data that rendered this figure are available at https://bitbucket.org/caseywdunn/animal-genomes (commit d40ba88).
Source: Photos by Casey Dunn.

that, while genomes were radically transformed along the branch that gave rise to animals, there is far less diversity within animals in terms of gene number and other genome features than expected [5,16–24,25••,26]. The genomes of animals that are sometimes erroneously referred to as 'higher,' such as vertebrates, are remarkably similar to those that were characterized as 'lower', such as sea anemones and sponges. This has been considered a paradox [27–29], but it says far more about our conceptions of animals and complexity than it does about the biology of the animals themselves. It is based on biases that lead us to greatly underestimate the intricacy of many poorly studied animals [30] and chauvinism that leads us to overestimate vertebrate complexity [28•], and presumes that there are a small number of common types of genome changes (such as acquisition of new protein

coding genes) that underlie particular phenotypic changes.

In order to understand the evolutionary history of animal genome diversity, it is helpful to first identify the basic features of the genome of the most recent common ancestor of all animals. By analyzing a broad sampling of genomes from animals and their close relatives in a phylogenetic context, it is possible to infer a minimal set of genes that were present in this organism. The current availability of genomic data from animals and animal outgroups is sparse, but does span the root of the animal tree (since genome sequences are available for both sponges and ctenophores, this holds regardless of which is the sister group to all other animals [16,17,31–34]). Based on the gene inventories described in [17], the most recent common ancestor of all animals had at least 6289 gene families that are also shared with outgroups and 2141 gene families that appear to be animal-specific. This number will be revised upward as taxon sampling improves, and may be revised downward if improved analysis methods detect previously overlooked homology. Though taxon sampling is still minimal, some interesting qualitative patterns have emerged regarding the composition of the genome of the most recent common ancestor of all animals (Table 1). For example, while many cell adhesion genes in animals are shared with outgroups, many cell signalling genes are animal-specific [19••].

The genome of the most recent common animal ancestor gave rise to the genomes of all living animals through many types of evolutionary changes. These include whole genome duplication events, changes in synteny [35•], gene family evolution (through the origin, duplication, and loss of genes) [36•], coding sequence evolution (through site insertion, deletion and substitution), modifications of regulatory regions, horizontal gene transfer [11], expansion and loss of transposable elements, and modification of local repeats (including satellite DNA, simple sequence repeats, and tandem repeats). The genomic features affected by these processes include introns [37], centromeres, telomeres, sex chromosomes, transposable elements, sequence-dependent variation in chromatin-organization, gene regulatory elements, as well as microRNAs and other noncoding RNAs.

Perhaps the most conspicuous of genome changes are changes in genome size [38]. Despite the relative stability in protein-coding gene content, animal genome sizes are highly variable and these changes are thought to be largely driven by nonadaptive processes [39]. In at least some clades, genome size correlates with cell size, cell division rate, resting metabolic rate, average adult size and therefore is potentially impacted by selection [40].

While gene inventory is more conserved across animals than had been expected, there are many genes that are found only in particular animal subclades [41]. These taxon-restricted 'orphan' genes may be formed de novo from noncoding sequence or alternatively may be highly derived genes that no longer show a signature of homology at the sequence level [42,43]. At least some of these genes have the ability to quickly become essential [44], but it is not yet clear if there are general patterns in the association of the origin of novel genes and phenotype evolution [41,45].

Connecting genome evolution to phenotype evolution

Not all changes in genomes result in changes in phenotypes — there is a 'many-to-one' mapping between genomes and phenotypes [46]. This has sweeping implications for some of the most basic questions in evolutionary genomics — What genomic differences across the animal phylogeny underlie the diversity of animal phenotypes? What is the phenotypic impact of particular genomic differences? What phenotypes are evolutionarily accessible? Because there are so many genomic differences and so many of them have no impact on phenotypes it is extremely difficult to make causative associations between the two. The most common strategy is to extrapolate experimental results from laboratory model organisms to other species of interest (i.e., the candidate gene approach), but this can introduce considerable bias since it is largely blind to novel features in nonmodel species [30]. Another approach is to collect enough comparisons between genotypes and phenotypes to make statistically significant associations between particular evolutionary changes in the genome and particular evolutionary changes in the phenotype [47•], whether at the population or macroevolutionary scale. This requires genomic data from a large number of organisms that are variable in the phenotypic trait of interest, and many independent evolutionary changes in the trait of interest. Even so, it has already proven tractable for specific traits in clades such as mammals [48] where taxon sampling is already quite good and there is much experimental evidence regarding gene function in some species. This approach will be more powerful as genome and phenotype sampling improve. Wider adoption of these approaches will also hinge on developing analysis methods that can account for between species variation as well as within speciation variation [49], and approaches that can extend phylogenetic comparative methods to high dimensional data [50]. The incorporation of functional genomic data sampled across a broad diversity of species, including gene expression data [50,51], will also greatly assist in understanding the phenotypic implications of genomic changes.

Identifying the evolutionary processes that shape genome evolution

Not all changes in phenotype result in a change in fitness — there is a 'many-to-one' mapping between phenotypes

Table 1

Taxon-specific gene families for clades shown in Figure 1

Clade	Gene	Reference
Metazoa	LIM classes (ABLIM, Enigma, LHX, LMO7, Mical)	[65]
Metazoa	HD: ANTP, PRD, POU, LIM, SINE classes	[17]
Metazoa	Sox (B, C, E, F)	[66]
Metazoa	Class I Fox TFs: D, G, L2	[36•]
Metazoa	bHLH: Group A (E12/E47, ProtoAtonal), E.	[36•]
Metazoa	bZIP: Jun, Fos, XBP1, MAF, Nfe2	[36•]
Metazoa	Ets, Smad, NR, Interferon Regulatory Factor, MADF, AP-2, Doublesex	[36•]
Metazoa	T-box: Eomes, Tbx2/3, Tbx8, TbxPor, Tbx1/15/20	[36•]
Metazoa	Wnt	[67]
Metazoa	LGR	[68]
Metazoa	Shaker K+ channels (Shaker)	[69]
Metazoa	EAG K+ channels	[70]
Porifera + Parahoxozoa	Immune genes: IRF, MyD88, TLR2	[71]
Porifera + Parahoxozoa	Axin	[67]
Porifera + Parahoxozoa	iGluR: (AMPA)	[71]
Porifera + Parahoxozoa	ADAR1	[72]
Porifera + Parahoxozoa	Nuclear receptors with zinc fingers	[73]
Porifera + Parahoxozoa	homeodomain + paired domain	[17]
Porifera + Parahoxozoa	Drosha, Pasha, microRNAs	[74]
Parahoxozoa	HD: HNF, PRD (S50 and K50), Hox	[17]
Parahoxozoa	Perlecan/HSPG2	[75]
Parahoxozoa	LIM: LMO class	[65]
Parahoxozoa	RHD: NFAT and Rel	[36•]
Parahoxozoa	GCM	[36•]
Parahoxozoa	Tbx6	[36•]
Parahoxozoa	Class I Fox TFs: A, B, C, E, Q1, Q2	[36•]
Parahoxozoa	bHLH: Group A (Achaete Scute, Paraxis, Twist, Net, Mesp), Group C (Bmal, Hairy), Group D	[36•]
Parahoxozoa	TGF-beta: Noggin, Follistatin	[76]
Parahoxozoa	Contactin, Neurexin	[77]
Parahoxozoa	BMP-antagonist	[68]
Parahoxozoa	Shaker K+ channels (Shab, Shaw)	[69]
Parahoxozoa	EAG K+ channels (Eag, Elk,Erg)	[70]
Parahoxozoa	Gustatory receptor-like	[78]
Planulozoa	Wnt: Wnt2, Wnt3, Wnt4, Wnt5, Wnt7, Wnt11, Wnt16, DKK	[67]
Planulozoa	Nuclear receptors: NR2F, NR1/4, NR2B, NR2C/D/E	[73]
Planulozoa	TGF-beta: Chordin	[76]
Planulozoa	Erbin, CASK, Neuroligin	[17]
Planulozoa	mir-100	[74]
Planulozoa	PDGF	[68]
Planulozoa	Shaker K+ channels (Shal)	[69]
Planulozoa	Cyclooxygenase	[79]
Bilateria	bZIP: BACH, B-ATF, Atf3	[36•]
Bilateria	ClassIFoxTFs: F, H,I,L1	[36•]
Bilateria	HD: CUT, PROS, ZF	[36•]
Bilateria	HMGbox: Sox-like genes	[66]
Bilateria	bHLH: Group A (MyoD, NeuroD, Neurogenin, MyoR), Group B (SRC)	[36•]
Bilateria	Stargazin	[17]

and fitness. Because genotype to phenotype mapping is also many-to one (as discussed above), it is extraordinarily difficult to understand whether differences in genome sequences result in differences in phenotypes that in turn result in fitness differences that can be acted on by natural selection. While there has been a historical bias to invoke adaptationist arguments for genome variation, others have powerfully argued that most genome variation is neutral (i.e., has no impact on fitness) [52].

There are several ways to investigate whether genome variation is the result of neutral evolution or has been shaped by natural selection. First, one can measure the impact of genotype variation on phenotype variation and phenotype variation on fitness. This is exceptionally difficult, though, as most of these steps are difficult and very labor intensive. Second, one can search directly for signatures of selection in the genome, such as selective sweeps, phylogenetic footprints, or deviations in the

nonsynonymous mutation rate [53], without investigating the phenotypes that mediated the selection [54].

Genome evolution as a proxy

Many evolutionary biologists that collect and analyze genome data are not primarily motivated by the desire to study genome evolution — they seek to use the evolutionary history of the genomes as a proxy to understand other aspects of evolution. Population geneticists use genomes to infer demographic events in the history of populations, such as migration and ancestral population sizes [55]. Phylogenetic systematics now uses genome-scale data to infer the relationships between species.

The first animal phylogenies based entirely on genome data [56] supported relationships, including the Coelomata hypothesis, that were later shown with the addition of more data and improved analysis methods to be artefacts [33,57]. This reinforces how important broad taxonomic sampling is for interpreting genome data. Even if an investigator is interested in taxonomically restricted questions, such as genome evolution in a particular clade, there is great value in considering those questions in the context of genome data from additional species. Most phylogenetic analyses based on genome data now also include transcriptome sampling [33,57], which can be a less expensive way to expand taxon sampling. The vast majority of studies that use genomic data to investigate phylogenetic relationships consider only the molecular sequence evolution of protein coding genes. There is also considerable interest and great promise in tapping other types of genomic data, such as gene gain and loss and noncoding sequences, to understand species relationships [58].

Conclusions

The real power in animal evolutionary genomics will not be in what we learn about genomes themselves, but how we connect this information to phenotypic diversity and the historical processes that led to the diversity of genomes we see today. As genome sequences become available from a greater number of species, phylogenetic comparative methods [49,59] will become increasingly important for interpreting the biological implications of genome diversity and leveraging the diversity of genomes across species to understand genome function within species. They will first need to be retooled to work effectively with high dimensional data, though. There are many phenotype changes and many more genome changes along each branch of the animal tree. Just because a genome change is evolutionarily coincident with a particular phenotype change does not mean it is causative. It could have no phenotypic impact at all, or be related to one of the many other phenotypic changes along the same phylogenetic branch. Unfortunately, such spurious associations are still a mainstay of animal genome papers that purport to identify genes specific to phenotypic novelties [60,61]. One of the most common problems with such conclusions is the failure to account for multiple independent tests.

The concept of 'a genome' for a given animal species is an invention of convenience — there is no such thing. Species are sets of individuals with many different genomes. Genomes can even vary between cells within individuals due to programmed genome rearrangement [62] and somatic mutation. As the field matures, many evolutionary questions will require sampling genomes at all these scales (multiple cells from multiple individuals of multiple species).

The field of evolutionary animal genomics needs to stop looking for relationships between vague notions of genome complexity and vague notions of phenotype complexity, and instead focus on understanding the evolution of genes and phenotypes on a trait-by-trait basis. There is no single axis of organism complexity, each animal has a mix of traits that could be characterized as simple or complex. In addition, our sense of differences in complexity across animals is inflated by biases that lead us to focus on complex traits in well-studied animals and miss complex novel features in poorly studied animals [30]. There is also reason to question the expected differences in genomes. Because there is a many-to-one mapping between genomes, phenotypes, and fitness, there are many different changes in genomes that can lead to equivalent phenotype changes. Instead of seeking a small number of genome evolutionary patterns that explain sweeping patterns in animal diversity, such as the gain of novel protein coding genes or micro-RNAs, we should be prepared to identify and explain very different genotype evolution stories that underlie every phenotypic change.

Acknowledgements

Thanks to Felipe Zapata and Stefan Siebert for providing feedback on this manuscript. This work was supported in part by the National Science Foundation Alan T Waterman Award to CWD.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Kay AC: *The Computer Revolution Hasn't Happened Yet [Internet]*. 1997.

2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**:860-921.

3. Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W *et al.*: **Evolution of genes and genomes on the *Drosophila* phylogeny**. *Nature* 2007, **450**:203-218.

4. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT *et al.*: **Whole-genome analyses resolve early branches in the tree of life of modern birds**. *Science* 2014, **346**:1320-1331.

5. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K *et al.*: **The amphioxus genome and the evolution of the chordate karyotype**. *Nature* 2008, **453**:1064-1071.

6. Flot J-F, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EGJ, Hejnol A, Henrissat B, Koszul R, Aury J-M *et al.*: **Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga***. *Nature* 2013, **500**:453-457.

7. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D-H, Larsson T, Lv J, Arendt D *et al.*: **Insights into bilaterian evolution from three spiralian genomes**. *Nature* 2014, **493**:526-531.

8. Wang B, Ekblom R, Bunikis I, Siitari H, Höglund J: **Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster-Z and MHC evolution**. *BMC Genomics* 2014, **15**:180.

9. Zhan S, Merlin C, Boore JL, Reppert SM: **The monarch butterfly genome yields insights into long-distance migration**. *Cell* 2011, **147**:1171-1185.

10. Richards S: **It's more than stamp collecting: how genome sequencing can unify biological research**. *Trends Genet* 2015 http://dx.doi.org/10.1016/j.tig.2015.04.007.

11. Artamonova II, Lappi T, Zudina L, Mushegian AR: **Prokaryotic genes in eukaryotic genome sequences: when to infer horizontal gene transfer and when to suspect an actual microbe**. *Environ Microbiol* 2015 http://dx.doi.org/10.1111/1462-2920.12854.

12. Kelley DR, Salzberg SL: **Detection and correction of false segmental duplications caused by genome mis-assembly**. *Genome Biol* 2010, **11**:R28.

13. Treangen TJ, Salzberg SL: **Repetitive DNA and next-generation sequencing: computational challenges and solutions**. *Nat Rev Genet* 2012, **13**:36-46.

14. • GIGA Community of Scientists: **The global invertebrate genomics alliance (GIGA): developing community resources to study diverse invertebrate genomes**. *J Hered* 2014, **105**:1-18.
This review highlights regions of the animal tree that are undersampled and provides detailed justification for generating genomic sequences from animals within these lineages.

15. Bird AP: **Gene number, noise reduction and biological complexity**. *Trends Genet* 1995, **11**:94-100.

16. Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM *et al.*: **The ctenophore genome and the evolutionary origins of neural systems**. *Nature* 2014, **510**:109-114.

17. Ryan JF, Pang K, Schnitzler CE, Nguyen AD, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, NISC Comparative Sequencing Program *et al.*: **The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution**. *Science* 2013, **342** 1242592–1242592.

18. Fairclough SR, Chen Z, Kramer E, Zeng Q, Young S, Robertson HM, Begovic E, Richter DJ, Russ C, Westbrook MJ *et al.*: **Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta***. *Genome Biol* 2013, **14**:R15.

19. •• Suga H, Chen Z, de Mendoza A, Sebé-Pedrós A, Brown MW, Kramer E, Carr M, Kerner P, Vervoort M, Nchez-Pons Nursa *et al.*: **The *Capsaspora* genome reveals a complex unicellular prehistory of animals**. *Nat Commun* 2013, **4**:1-9.
This genome paper greatly clarified the understanding of genome evolution in Holozoa (the clade that includes animals) and provided important insight into the changes in genome composition that preceded the radiation of animals.

20. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, Fujie M, Fujiwara M, Koyanagi R, Ikuta T *et al.*: **Using the *Acropora digitifera* genome to understand coral responses to environmental change**. *Nature* 2011, **476**:320-323.

21. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U *et al.*: **The *Amphimedon queenslandica* genome and the evolution of animal complexity**. *Nature* 2010, **466**:720-726.

22. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D *et al.*: **The dynamic genome of *Hydra***. *Nature* 2010, **464**:592-596.

23. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML *et al.*: **The *Trichoplax* genome and the nature of placozoans**. *Nature* 2008, **454**:955-960.

24. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I *et al.*: **The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans**. *Nature* 2008, **451**:783-788.

25. •• Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV *et al.*: **Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization**. *Science* 2007, **317**:86-94.
This is the first animal genome to be sequenced outside of Bilateria. It revealed that many genome features were much older than previously thought, and that many of the differences between vertebrate and ecdysozoan genomes were due to reductions in ecdysozoa, not expansions in vertebrates.

26. Sea Urchin Genome Sequencing Consortium, Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, Angerer LM, Arnone MI, Burgess DR *et al.*: **The genome of the sea urchin *Strongylocentrotus purpuratus***. *Science* 2006, **314**:941-952.

27. Copley RR: **The animal in the genome: comparative genomics and evolution**. In *Animal evolution: genomes, fossils, and trees.* Edited by Telford , Littlewood DTJJ. Oxford University Press; 2009: 148-156.

28. • Hahn MW, Wray GA: **The g-value paradox**. *Evol Dev* 2002, **4**:73-75.
A concise summary of the paradox, or lack thereof, between the number of protein coding genes and morphological complexity.

29. Taft RJ, Pheasant M, Mattick JS: **The relationship between non-protein-coding DNA and eukaryotic complexity**. *Bioessays* 2007, **29**:288-299.

30. Dunn CW, Leys SP, Haddock SHD: **The hidden biology of sponges and ctenophores**. *Trends Ecol Evol* 2015, **30**:282-291.

31. Dunn CW, Giribet G, Edgecombe GD, Hejnol A: **Animal phylogeny and its evolutionary implications**. *Annu Rev Ecol Evol Syst* 2014, **45**:371-395.

32. Whelan NV, Kocot KM, Moroz LL, Halanych KM: **Error, signal, and the placement of Ctenophora sister to all other animals**. *Proc Natl Acad Sci* 2015 http://dx.doi.org/10.1073/pnas.1503453112.

33. Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD *et al.*: **Broad phylogenomic sampling improves resolution of the animal tree of life**. *Nature* 2008, **452**:745-749.

34. Nosenko T, Schreiber F, Adamska M, Adamski M, Eitel M, Hammel J, Maldonado M, Müller WEG, Nickel M, Schierwater B *et al.*: **Deep metazoan phylogeny: when different genes tell different stories**. *Mol Phylogenet Evol* 2013, **67**:223-233.

35. • Srivastava M: **A comparative genomics perspective on the origin of multicellularity and early animal evolution**. *Evolutionary transitions to multicellular life*. Netherlands: Springer; 2015, 269-299.
A detailed recent review on comparative animal genomics.

36. • Sebé-Pedrós A, de Mendoza A: **Transcription factors and the origin of animal multicellularity**. In *Evolutionary transitions to multicellular life.* Edited by Ruiz-Trillo I, Nedelcu A. 2015:379-394.
A detailed review of many papers to summarize the phylogenetic distribution of animal transcription factors. The work provides insight into evolutionary forces that shaped the repertoire of transcription factors in animals, how gene regulatory networks influenced the evolution of transcription factors, and what forces led to the extraordinary regulatory capabilities of animal transcription factors.

37. Roy S, Gilbert W: **Resolution of a deep animal divergence by the pattern of intron conservation**. *Proc Natl Acad Sci* 2005, **102**:4403-4408.

38. Gregory TR: *The evolution of the genome*. Academic Press; 2011.

39. Lynch M, Conery JS: **The origins of genome complexity**. *Science* 2003, **302**:1401-1404.

40. Gregory TR: **Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma**. *Biol Rev Cambridge Philos Soc* 2001, **76**:65-101.

41. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG: **More than just orphans: are taxonomically-restricted genes important in evolution?** *Trends Genet* 2009, **25**:404-413.

42. Tautz D, Domazet-Lošo T: **The evolutionary origin of orphan genes**. *Nat Rev Genet* 2011, **12**:692-702.

43. Ding Y, Zhou Q, Wang W: **Origins of new genes and evolution of their novel functions**. *Annu Rev Ecol Evol Syst* 2012, **43**:345-363.

44. Chen S, Zhang YE, Long M: **New genes in *Drosophila* quickly become essential**. *Science* 2010, **330**:1682-1685.

45. Forêt S, Knack B, Houliston E, Momose T, Manuel M, innec EQ, Hayward DC, Ball EE, Miller DJ: **New tricks with old genes: the genetic bases of novel cnidarian traits**. *Trends Genet* 2010 http://dx.doi.org/10.1016/j.tig.2010.01.003.

46. Stadler BMR, Stadler PF, Wagner GP, Fontana W: **The topology of the possible: formal spaces underlying patterns of evolutionary change**. *J Theor Biol* 2001, **213**:241-274.

47. • Wray GA: **Genomics and the evolution of phenotypic traits**. *Annu Rev Ecol Evol Syst* 2013, **44**:51-72.
Review shows how genomic approaches are being applied to long-standing issues in evolutionary genetics and how these approaches are expanding the field.

48. Hiller M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, Bejerano G: **A 'forward genomics' approach links genotype to phenotype using independent phenotypic losses among related species**. *Cell Rep* 2012, **2**:817-823.

49. Felsenstein J: **Comparative methods with sampling error and within-species variation: contrasts revisited and revised**. *Am Nat* 2008, **171**:713-725.

50. Dunn CW, Luo X, Wu Z: **Phylogenetic analysis of gene expression**. *Integr Comp Biol* 2013, **53**:847-856.

51. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M *et al.*: **The evolution of gene expression levels in mammalian organs**. *Nature* 2011, **478**:343-348.

52. Lynch M: *The origins of genome architecture*. Sinauer Associates Incorporated; 2007.

53. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution**. *Mol Biol Evol* 1998, **15**:568-573.

54. Nielsen R: **Molecular signatures of natural selection**. *Annu Rev Genet* 2005, **39**:197-218.

55. Li H, Durbin R: **Inference of human population history from individual whole-genome sequences**. *Nature* 2011 http://dx.doi.org/10.1038/nature10231.

56. Wolf YI, Rogozin IB, Koonin EV: **Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis**. *Genome Res* 2004, **14**:29-36.

57. Philippe H, Lartillot N, Brinkmann H: **Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia**. *Mol Biol Evol* 2005, **22**:1246-1253.

58. Telford MJ, Copley RR: **Improving animal phylogenies with genomic data**. *Trends Genet* 2011, **27**:186-195.

59. Felsenstein J: **Phylogenies and the comparative method [Internet]**. *Am Nat* 1985, **125**:1-15.

60. Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, Wynne JW, Xiong Z, Baker ML, Zhao W *et al.*: **Comparative analysis of bat genomes provides insight into the evolution of flight and immunity**. *Science* 2013, **339**:456-460.

61. Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, Robertson HM, Hudson ME, Venkat A, Fischman BJ *et al.*: **Genomic signatures of evolutionary transitions from solitary to group living**. *Science* 2015 http://dx.doi.org/10.1126/science.aaa4788.

62. Smith JJ, Baker C, Eichler EE, Amemiya CT: **Genetic consequences of programmed genome rearrangement**. *Curr Biol* 2012, **22**:1524-1529.

63. Appeltans W, Ahyong ST, Anderson G, Angel MV, Artois T, Bailly N, Bamber R, Barber A, Bartsch I, Berta A *et al.*: **The magnitude of global marine species diversity**. *Curr Biol* 2012, **22**:2189-2202.

64. Hamilton AJ, Basset Y, Benke KK, Grimbacher PS, Miller SE, Novotný V, Samuelson GA, Stork NE, Weiblen GD, Yen JDL: **Quantifying uncertainty in estimation of tropical arthropod species richness**. *Am Nat* 2010, **176**:90-95.

65. Koch BJ, Ryan JF, Baxevanis AD: **The diversification of the LIM superclass at the base of the metazoa increased subcellular complexity and promoted multicellular specialization**. *PLoS ONE* 2012, **7**:e33261.

66. Schnitzler CE, Simmons DK, Pang K, Martindale MQ, Baxevanis AD: **Expression of multiple Sox genes through embryonic development in the ctenophore *Mnemiopsis leidyi* is spatially restricted to zones of cell proliferation**. *EvoDevo* 2014, **5**:15-17.

67. Pang K, Ryan JF, Program NCS, Mullikin JC, Baxevanis AD, Martindale MQ: **Genomic insights into Wnt signaling in an early diverging metazoan, the ctenophore *Mnemiopsis leidyi***. *EvoDevo* 2010, **1**:10.

68. Roch GJ, Sherwood NM: **Glycoprotein hormones and their receptors emerged at the origin of metazoans**. *Genome Biol Evol* 2014, **6**:1466-1479.

69. Li X, Liu H, Chu Luo J, Rhodes SA, Trigg LM, van Rossum DB, Anishkin A, Diatta FH, Sassic JK, Simmons DK *et al.*: **Major diversification of voltage-gated K+ channels occurred in ancestral parahoxozoans**. *Proc Natl Acad Sci* 2015, **112** E1010–9.

70. Li X, Martinson AS, Layden MJ, Diatta FH, Sberna AP, Simmons DK, Martindale MQ, Jegla TJ: **Ether-à-go-go family voltage-gated K+ channels evolved in an ancestral metazoan and functionally diversified in a cnidarian-bilaterian ancestor**. *J Exp Biol* 2015, **218**:526-536.

71. Riesgo A, Farrar N, Windsor PJ, Giribet G, Leys SP: **The analysis of eight transcriptomes from all Poriferan classes reveals surprising genetic complexity in sponges**. *Mol Biol Evol* 2014, **31**:1102-1120.

72. Grice LF, Degnan BM: **The origin of the ADAR gene family and animal RNA editing**. *BMC Evol Biol* 2015, **15**:4.

73. Reitzel AM, Pang K, Ryan JF, Mullikin JC, Martindale MQ, Baxevanis AD, Tarrant AM: **Nuclear receptors from the ctenophore *Mnemiopsis leidyi* lack a zinc-finger DNA-binding domain: lineage-specific loss or ancestral condition in the emergence of the nuclear receptor superfamily?** *EvoDevo* 2011, **2**:3.

74. Maxwell EK, Ryan JF, Schnitzler CE, Browne WE, Baxevanis AD: **MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore *Mnemiopsis leidyi***. *BMC Genom* 2012, **13**:714.

75. Warren CR, Kassir E, Spurlin J, Martinez J, Putnam NH, Farach-Carson MC: **Evolution of the Perlecan/HSPG2 gene and its activation in regenerating *Nematostella vectensis***. *PLOS ONE* 2015, **10**:e0124578.

76. Pang K, Ryan JF, Baxevanis AD, Martindale MQ: **Evolution of the TGF-β signaling pathway and its potential role in the ctenophore, *Mnemiopsis leidyi***. *PLoS ONE* 2011, **6**:e24152.

77. Ganot P, Zoccola D, Tambutté E, Voolstra CR, Aranda M, Allemand D, Tambutté S: **Structural molecular components of**

septate junctions in cnidarians point to the origin of epithelial junctions in eukaryotes. *Mol Biol Evol* 2015, **32**:44-62.

78. Saina M, Busengdal H, Sinigaglia C, Petrone L, Oliveri P, Rentzsch F, Benton R: **A cnidarian homologue of an insect gustatory receptor functions in developmental body patterning**. *Nat Commun* 2015, **6**:6243.

79. Havird JC, Kocot KM, Brannock PM, Cannon JT, Waits DS, Weese DA, Santos SR, Halanych KM: **Reconstruction of cyclooxygenase evolution in animals suggests variable, lineage-specific duplications, and homologs with low sequence identity**. *J Mol Evol* 2015, **80**:193-208.