Genome **Biology**

Sequencing the genome of the Burmese python (*Python molurus bivittatus*) as a model for studying extreme adaptations in snakes

Todd A Castoe[1], AP Jason de Koning[1], Kathryn T Hall[1], Ken D Yokoyama[1], Wanjun Gu[2], Eric N Smith[3], Cédric Feschotte[3], Peter Uetz[4], David A Ray[5], Jason Dobry[6], Robert Bogden[6], Stephen P Mackessy[7], Anne M Bronikowski[8], Wesley C Warren[9], Stephen M Secor[10] and David D Pollock[1]*

Abstract

The Consortium for Snake Genomics is in the process of sequencing the genome and creating transcriptomic resources for the Burmese python. Here, we describe how this will be done, what analyses this work will include, and provide a timeline.

The importance of snakes, and the Burmese python, as model organisms

The evolutionary origin of snakes involved extensive morphological and physiological adaptations that included the loss of limbs, lung reduction, and trunk and organ elongation. Most snakes also evolved a suite of radical adaptations to consume large prey relative to their body size, including the ability to endure extreme physiological and metabolic fluctuations [1,2] and produce diverse venom proteins [3,4]. These radical adaptations, centered around consuming large prey whole, have made snakes an interesting model for studying metabolic flux and organ physiology, regeneration, and regulation, with the most important example being the Burmese python.

Within 2 to 3 days after feeding, the Burmese python (*Python molurus bivittatus*) can experience tremendous physiological changes, including: a 44-fold increase in metabolic rate (the highest among tetrapods); 35 to 100% increases in the mass of the heart, liver, pancreas, small intestine, and kidneys; 160-fold increase in plasma fatty acid and triglyceride content; and 5-fold increase in intestinal microvillus length [1,5]. After the completion of digestion, each of these phenotypes is reversed as digestive functions are downregulated and tissues

*Correspondence: David.Pollock@ucdenver.edu
[1]Department of Biochemistry & Molecular Genetics, University of Colorado School of Medicine, 12801 17th Ave, Aurora, CO 80045, USA
Full list of author information is available at the end of the article

undergo atrophy [6]. This extreme modulation of tissue morphology and function facilitates investigation into the signaling and cellular mechanisms that underlie regulation of organ performance and regeneration. These animals are also readily obtained from commercial breeders, non-aggressive, and easier and cheaper to care for than laboratory rats. The scientific potential of this system to reveal molecular mechanisms associated with these extreme reactions (and their reversal) is tremendous, and can provide novel insight into vertebrate gene and systems function, novel strategies and drug targets for treating human diseases, and alternative disease models.

Snakes have also been used as model species for high-profile discoveries pertaining to vertebrate development, including the findings that vertebrate metamerism (somitogenesis) can be controlled by changing the rate of somitogenesis [7], that the loss of limbs correlates with changes in expression of some regulatory genes [8] as well as *Hox* gene expression and gene structure [9], that particular developmental pathways are associated with tooth and fang development [10], and that limblessness in snakes may result from failure to activate core vertebrate signaling pathways during development and from changes in *Hox* gene expression [8,11]. Snakes are also important models for high-performance muscle physiology [12], genetic sex determination [13], evolutionary ecology [14,15], and molecular evolution and adaptation [16-18]. Enhanced snake genomic resources (eventually including comparative genomic data from multiple species) are expected to provide additional insight into how the unique structures and developmental processes of snakes evolved.

In addition to the python (which is non-venomous), venomous snake species are also important for biomedical research, as is developing a greater understanding of the genomic and adaptive contexts leading to the origin of venom genes. Worldwide, the World Health

Organization estimates that there are about 2.5 million venomous snake bites per year (about 1,400 in the US), resulting in about 125,000 deaths [19]. As a consequence, the health relevance of snake venom research is extensive. Genes identified in snake venoms are related to genes used in normal housekeeping and digestive roles in other vertebrates [3,4], but the details of how these have been modified by evolution to become functionally diverse toxic venoms cannot readily be determined without good comparative information from the full complement of genes from both venomous and non-venomous snakes.

Phylogenetic position of snakes and the python

Among vertebrates, the snake lineage represents a speciose (about 3,100 species) and phenotypically diverse radiation. Because snakes represent such an ancient (about 150 million years old) lineage on the branch of the vertebrate tree of life (Figure 1; squamate reptile divergence estimates based on [20]), understanding the content of snake genomes will contribute broadly to an understanding of vertebrate genomics. Together with the genome of the *Anolis* lizard, the availability of a snake genome (and eventually, multiple snake genomes) will contribute to better rooting of mammalian gene trees, and to more accurate reconstructions of amniote ancestral genome attributes. Below, we outline that in addition to the python genome, the genomes of the venomous king cobra and the non-venomous garter snake are also currently being sequenced. In the phylogenetic tree in Figure 1, we highlight that in addition to the major lineages being targeted by these three confirmed genome projects, there are two other major groups, blindsnakes and venomous vipers (for example, rattlesnakes), that are not yet explicitly targeted by ongoing genome sequencing projects (although multiple groups have cited these as potential targets). One purpose of the website that we have established [21] is to provide the community with updated information on targeting of species for genome sequencing.

Python genome project overview

A main goal of the python genome project is to provide key genomic resources to facilitate studies of how its extreme phenotypes are regulated and accomplished at the molecular level. Thus, a central component of the python genome project is to produce a draft python genome that contains genic and near-genic regions that are assembled and annotated. To provide a service to the broader research community, we have released a pre-publication preliminary draft assembly of the python genome for conditional use. We are working under the Toronto Statement for prepublication release [22], and this letter provides the details of our plans and responsibilities, as outlined in the original paper describing this statement [22].

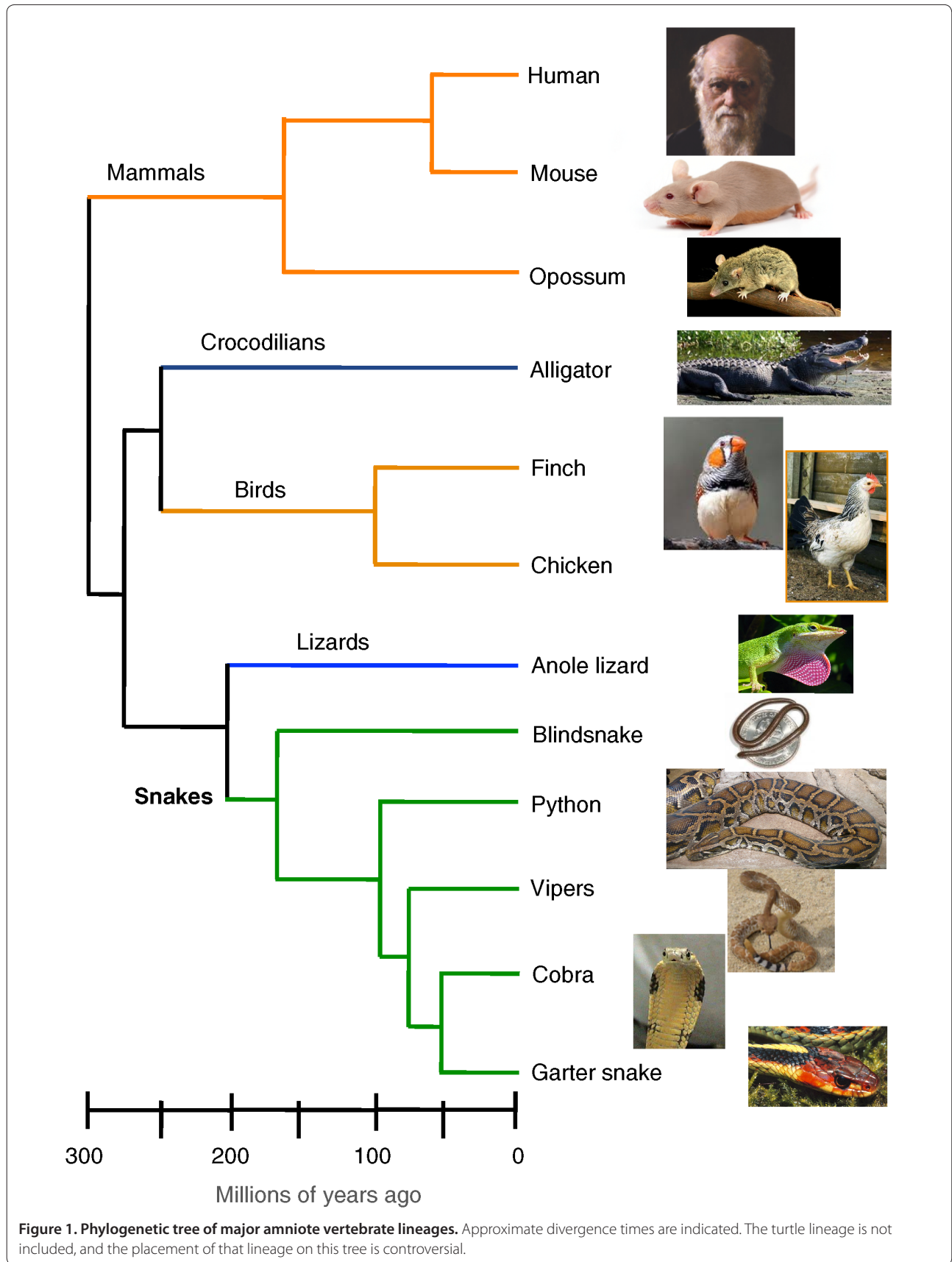### Properties of the python genome, and genomic resources currently available

Snake genomes are often smaller than mammalian genomes, ranging from about 1.3 Gbp to 3.8 Gbp, with an average of 2.08 Gbp [23]. There is no existing estimate for the genome of *Python molurus*, but the most recent estimate for the related species *Python reticulatus* is 1.44 Gbp; this suggests that the Burmese python genome is relatively small compared with most snakes. The karyotype of the Burmese python is known, and comprises 36 chromosomes (2n = 36), with 16 macrochromosomes and 20 microchromosomes [24]. All snakes are thought to have ZW genetic sex determination, with males being the homogametic sex (ZZ) and females heterogametic (ZW).
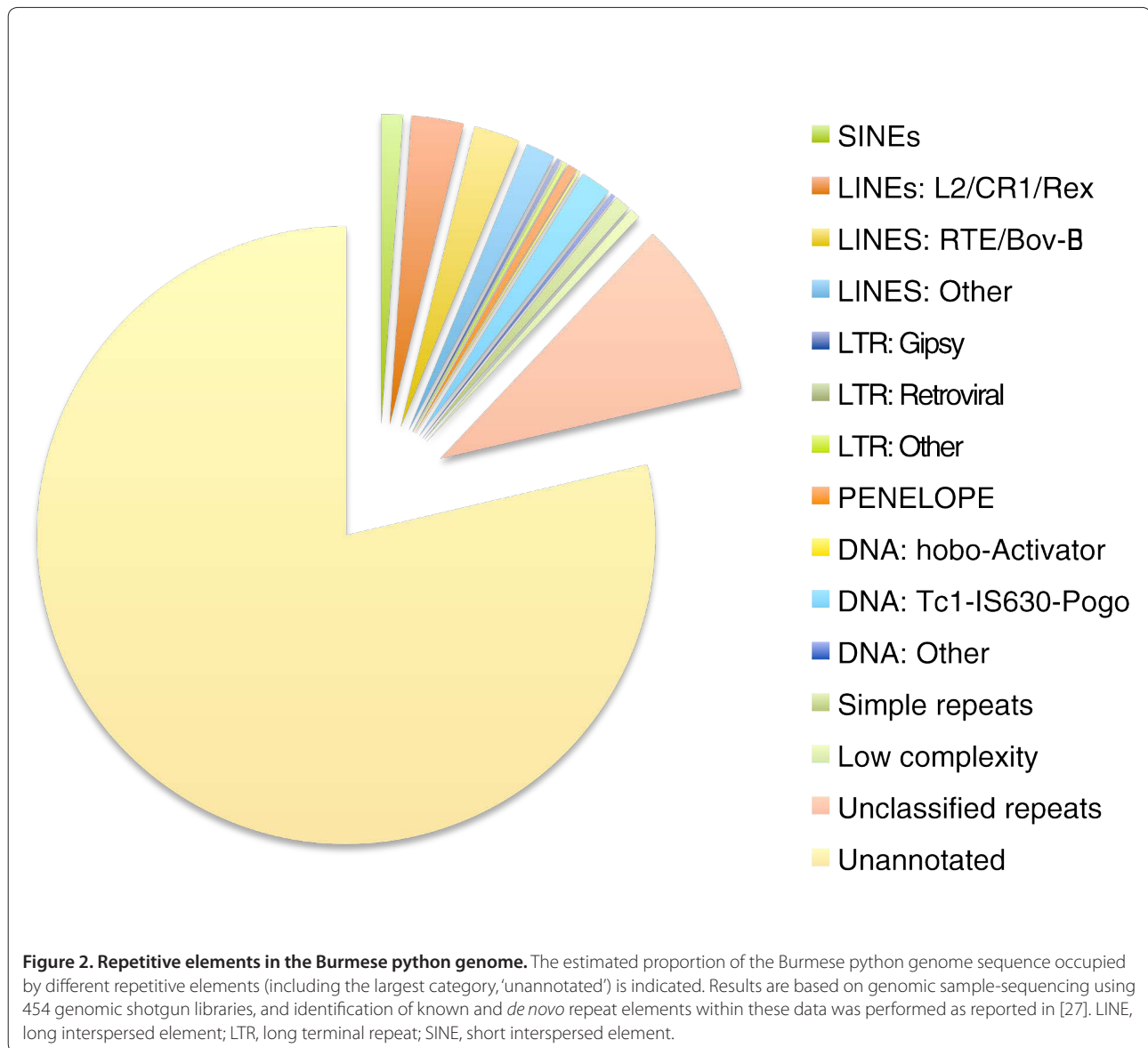
Since the early work of Olmo and colleagues [25,26] using DNA reassociation kinetics, it has been known that the genome of *P. molurus* had particularly low amounts of repetitive DNA compared with other snakes. This was recently confirmed with sequence-based evidence [27], using 454 sequencing of genomic shotgun libraries to randomly sample fractions of snake genomes, and using these fractions to estimate genomic repetitive element content and diversity (Figure 2; data based on [27]). From these data, the python genome was estimated to be made up of 21% readily identifiable repetitive element sequence (Figure 2), compared with more than double that (45%) in the venomous copperhead (a relative of the rattlesnake) with a similarly sized genome [27]. Despite the contrast in repetitive element abundance, both snakes contained a similarly broad diversity of transposable element types, which seems to be an emerging hallmark of squamate reptile (lizards and snakes) genomes [27-29]. Bov-B and CR1 LINE retroelements were among the most prominent transposable element types in the python genome (Figure 2) [27], a characteristic in common with other snake genomes [27,29].

Burmese python genome draft version 1.0

We completed and publicly released an initial draft assembly of the Burmese python genome (v1.0). This sequence was obtained from a single individual purchased from a commercial breeder, and did not originate from an inbred line (*per se*), and thus we expect moderate levels of heterozygosity.

This genome draft was built primarily from Illumina GAIIx sequencing of a short insert (325 bp) paired-end shotgun genomic library. Various amounts of sequence data were collected from this library using paired reads of three different lengths (114 bp: 15.1 Gbp, 76 bp: 5.6 Gbp, and 36 bp: 2.9 Gbp), with the addition of a small amount (30 Mbp) of 454 shotgun library sequences. The v1.0 draft Burmese python genome, based on 23.7 Gbp of DNA sequence data, is equivalent to approximately

Figure 1. Phylogenetic tree of major amniote vertebrate lineages. Approximate divergence times are indicated. The turtle lineage is not included, and the placement of that lineage on this tree is controversial.

Figure 2. Repetitive elements in the Burmese python genome. The estimated proportion of the Burmese python genome sequence occupied by different repetitive elements (including the largest category, 'unannotated') is indicated. Results are based on genomic sample-sequencing using 454 genomic shotgun libraries, and identification of known and *de novo* repeat elements within these data was performed as reported in [27]. LINE, long interspersed element; LTR, long terminal repeat; SINE, short interspersed element.

17-fold coverage of the estimated 1.4 Gbp python genome, and is available from the NCBI accession AEQU000000000.1. This coverage is equivalent to about 35X 'virtual' or 'structural' coverage of the genome, which includes the gaps in the paired-end sequences.

Computational genome assembly was conducted using SOAP *de novo* v.1.04, with a k-mer size of 31. This assembly yielded 1.128 million contigs, with a mean length of 944 bp and an N50 length of 1,355 bp. Using paired-end sequence reads, contigs were assembled into 324,418 scaffolds that had a mean length of 1,397 bp and an N50 length of 2,186 bp. The total length of the scaffolded assembly was 1,177 Mbp. We note that the average contig and scaffold sizes in this draft are relatively small, in part because there are no sequences from longer mate-pair libraries or BAC references to increase structural coverage and improve assembly; such coverage will be added in future drafts.

Python BAC library resources

There is a high-quality high-density (about 5X coverage) BAC library available for the Burmese python, constructed using DNA from the same individual from which the draft genome was sequenced. This BAC library, along with mapping and sequencing services, is currently available commercially to the public from Amplicon Express [30].

Other resources

Limited transcriptomic resources have already been made available at the snake genomics website [21], and a

larger suite of transcriptomic resources will be made available with the release of the second assembly of the python genome (v2.0). There is also a preliminary set of repeat element consensus sequences, estimated from genomic sample sequencing of 454 genomic shotgun libraries [21,27].

Strategy for sequencing the python genome

Our strategy for improving the existing python genome is to add substantial additional sequence coverage from slightly longer insert (600 bp) paired-end Illumina sequencing, together with 3-kb mate-pair paired-end sequence. We plan to have a total of 50X coverage of these mixed read types, predominantly from long (114 to 150 bp) Illumina GAIIx paired-end reads.

The second draft assembly will be updated with the new short and long insert paired-end sequence data. Genome assembly will involve four principal steps that progress from forming contigs from raw quality-filtered sequence reads, to connecting contigs into scaffolds using paired-end sequence data, to gap filling (using all reads) and error correction. The set of smaller contigs will serve as anchors for addition of longer range insert sizes to increase scaffold length.

We therefore expect that contig lengths will be sufficient for most gene predictions and post-assembly alignment-based analysis. We also expect that the attributes of the python genome, being smaller and also lower in repetitive content than mammalian genomes (or other snakes), for example [27], together with our use of relatively long sequence reads, will produce a reasonably good quality assembly with moderately long contigs and scaffolds.

We will assess the accuracy of the assembled python genome using several methods, including read chaff rate (proportion of reads not incorporated into the assembly), read depth of coverage, average quality values per contig, discordant read pairs, gene footprint coverage (as assessed by cDNA contigs) and comparative alignments to the most closely related species with a complete genome - the *Anolis* lizard (and eventually, other snake genome assemblies). We will also take advantage of mapped cDNA contigs from various python tissues to improve assembly contiguity and accuracy, further strengthening the genic component of this assembly.

Our internally contamination-screened genome assembly will be submitted to the whole genome shotgun division of GenBank for independent contamination analysis. The final assembly will be posted on the Ensembl [31], University of California Santa Cruz [32] and NCBI [33] genome browsers for public queries as soon as it is available and passes contamination analyses, and relevant announcements and links will be posted on the snake genomics website [21].

Description of sequencing project with anticipated milestones and timeline

We recently released a preliminary draft assembly of the python genome (v1.0) to the public, together with limited transcriptome data. This assembly includes primarily about 17X coverage from Illumina short-insert paired-end sequencing and is therefore expected to be relatively fragmentary. Our anticipated timeline includes the completion of data collection required for the updated assembly (v2.0) based on extended genome coverage (about 50X) from short and longer insert paired-end Illumina sequencing by the end of the summer of 2011. This will be accompanied by an extensive set of transcriptome data, from multiple organs, that will be incorporated into gene prediction annotations. Attainment of 50X genome coverage and completion of long mate-pair library sequencing will mark the end of the data collection phase and the start of assembly and analysis. The end of this phase will be marked clearly on the snake genomics website [21], as will milestones of data analysis and release. The maximum time between the end of data collection and submission of the genome paper will be 1 year. The Toronto Statement suggests that there be a 1-year period, after which global analyses and publication by the community would be unimpeded. We recognize the start of this 1-year period at approximately the time that this manuscript will be published, July 2011, and therefore this embargo period would end July 2012.

Biological questions and types of analyses to be addressed by the python genome project

Here we outline the major questions, types of analyses, and analytical goals that will be included in the core python genome marker paper. The Toronto Statement suggests this be done to identify these topics as being somewhat embargoed, and we also see this as providing expectations for the community regarding the types of analyses planned. Although vignettes of the topics below will, in most cases, appear in some form in the core python genome paper, a majority of these will also involve longer-term research (including other publications) by members of the working group. Ultimately, the goal of the Consortium for Snake Genomics is to make certain that research efforts are not duplicated, and also to put together clusters of researchers interested in similar questions. Thus, we continue to welcome additional members to join the Consortium for Snake Genomics, and because of this, the research scope of the group may continue to expand beyond even what we outline here because of the interests of new members.

The analytical goals of the python genome project focus on aspects of the extreme physiology and metabolism of pythons, and on making links between the extreme phenotypes and genotypes of the python and snakes in

general. A main focus of analysis will include transcriptome data that describes the dynamics of gene expression that accompanies major physiological transitions brought about by feeding in the python. We will also be conducting genome-wide analysis of protein evolution to detect patterns of molecular evolution indicating positive selection that may relate to key adaptations of snakes, and the python specifically. In addition to focusing on all proteins in the genome, we intend to include detailed analysis of sets of genes involved in physiology, metabolism, heat sensing, vision, body elongation, limb loss, and the evolution of snake venoms. We anticipate analyzing how the protein families of interest identified above have differentially expanded or contracted in the snake and mammalian lineages.

We are also interested in analyses that focus on areas of the genome outside of the protein-coding regions. Complementing our analysis of protein-coding genes, we plan to use the python genome to investigate, essentially for the first time, unique properties of snake and reptilian gene and promoter architecture, and to make a first attempt to identify snake *cis*-regulatory elements and compare these to other species. Specifically, this analysis will include comparisons of nucleotide content and over-represented motifs that occur in core upstream promoters of genes with well-predicted transcription starts. Our comparisons would highlight *cis*-regulatory structure in the python and anole lizard in relation to patterns in other vertebrates. We also are interested in studying the repetitive element landscape of the python genome, including identification of which types of transposable elements occur in the python genome and how these elements have expanded over evolutionary time, and how horizontal transfer may explain their origins in the python genome. Our genome analyses will additionally include identification of single nucleotide polymorphisms from genomic and transcriptomic data collected, and an effort to make available sets of sequences for use as molecular markers for snakes (for example, microsatellite primers and orthologous loci for use in phylogenetics and other applications). Lastly, we will be conducting a detailed analysis to identify genomic sequences that represent python sex chromosomes by using genomic sequences collected from multiple individuals from both sexes.

There are a number of potential research areas that would probably be productive to pursue but are outside of the scope of the current plans of the project - these topics are therefore potential research avenues that we encourage others to pursue. Because the python represents a relatively deep evolutionary lineage on the amniote vertebrate tree of life, using the python data together with other comparative data to estimate genomic characteristics of the ancestral amniote genome (or the ancestral squamate genome) would be fascinating, including estimation of ancestral gene family copy numbers, instances of differential expansion/contraction of gene families in mammals and squamate reptiles, evolution of long conserved non-coding sequences, and genomic features such as isochore structure. Analysis of genes and gene families involved in vertebrate hearing, locomotion, behavior, and coloration are other examples of projects outside of the scope of the current project.

Justification and strategies for expansion of snake genomics

Research incorporating snakes as model systems is becoming increasingly popular and diverse in its breadth of topics. The availability of the python genome and associated resources will provide a much-needed genetic and genomic reference infrastructure for further facilitating such research. In addition to the importance of the python as a model for research, different snake species have been used as model systems for different types of research. For example, research focusing on behavior, development, and evolutionary ecology has focused on smaller non-venomous species such as garter and corn snakes in the family Colubridae, whereas research related to snake venom and envenomation have centered on venomous species typically in the families Viperidae (for example, rattlesnakes, and adders) and Elapidae (for example, coral snakes, cobras, and mambas). In addition to these lineages that contain commonly used model research species, blindsnakes represent a lineage that diverged long ago from the rest of the snakes, and as such would be a major contribution for comparative and evolutionary analyses. In addition to the python, we are aware of two additional confirmed snake genome sequencing projects targeting the non-venomous garter snake [29], and the venomous king cobra (F Vonk, personal communication; Figure 1). We therefore expect that multiple snake genomes will be available to support diverse research projects in the near future, and the incorporation of additional lineages of snakes would further support their utility as research models.

Formation of the Consortium for Snake Genomics and a portal for snake genomic resources

To foster the growth of a productive and interactive community of researchers interested in snake genomics, and to also encourage the growth of snake genomic resources, we have established the Consortium for Snake Genomics (CSG) and a website to house related content [21]. A core concept guiding the establishment of the CSG is that through shared interest in developing resources for snake-related research, individual researchers would be able to benefit from the pooling of resources, research motivations, and expertise, while also avoiding redundant

effort. Therefore, an integral part of this vision includes the recruitment of, and interaction among, a diverse working group of researchers interested in using snake genomic resources.

The CSG is also directly involved with the reptilian subset of the Genome10K project [34], with the intention of making certain that efforts to build resources for particular species are not duplicated, and that scientific arguments for the need for genomic resources of particular types, or for particular snake lineages, get translated into priorities for future sequencing initiatives, and that all this gets translated to the community through the snake genomics website [21]. At the website we have created pages with links to available snake genomic resources, and posted updates (news) on major projects, such as the status of various snake genomics sequencing projects and data releases; RSS feeds have been set up so that changes to the various pages can be updated through RSS readers automatically once subscribed to the feed. We have also set up an email list system so that interested researchers can request to receive occasional email updates related to snake genomics. Lastly, for researchers interested in becoming directly integrated into ongoing or future CSG projects, email contacts for the lead author are provided on the site.

Author details

[1]Department of Biochemistry & Molecular Genetics, University of Colorado School of Medicine, 12801 17th Ave, Aurora, CO 80045, USA. [2]Key Laboratory of Child Development and Learning Science, Southeast University, Si Pai Lou 2, Ministry of Education, Nanjing, 210096, China. [3]Department of Biology, University of Texas, 501 S. Nedderman Dr., Arlington, TX 76019, USA. [4]Center for Bioinformatics & Computational Biology, University of Delaware, 15 Innovation Way, Newark, DE 19711, USA. [5]Department of Biochemistry and Molecular Biology, Mississippi State University, 101 College Road, Mississippi State, MS 39762, USA. [6]Amplicon Express, 2345 NE Hopkins Ct., Pullman, WA 99163, USA. [7]School of Biological Sciences, 501 20th Street, University of Northern Colorado, Greeley, CO 80631, USA. [8]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, 253 Bessey Hall, Ames, IA 50011, USA. [9]Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Ave, St Louis, MO 63108, USA. [10]Department of Biological Sciences, University of Alabama, 300 Hackberry Lane, Tuscaloosa, AL 35487, USA.

References

1. Secor SM, Diamond J: **Adaptive responses to feeding in Burmese pythons: pay before pumping.** *J Exp Biol* 1995, **198**:1313-1325.
2. Secor SM, Diamond J: **A vertebrate model of extreme physiological regulation.** *Nature* 1998, **395**:659-662.
3. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E: **Early evolution of the venom system in lizards and snakes.** *Nature* 2006, **439**:584-588.
4. Ikeda N, Chijiwa T, Matsubara K, Oda-Ueda N, Hattori S, Matsuda Y, Ohno M: **Unique structural characteristics and evolution of a cluster of venom phospholipase A: isozyme genes of *Protobothrops flavoviridis* snake.** *Gene* 2010, **461**:15-25.
5. Cox CL, Secor SM: **Matched regulation of gastrointestinal performance in the Burmese python, *Python molurus*.** *J Exp Biol* 2008, **211**:1131-1140.
6. Secor SM: **Digestive physiology of the Burmese python: broad regulation of integrated performance.** *J Exp Biol* 2008, **211**:3767-3774.
7. Gomez C, Ozbudak EM, Wunderlich J, Baumann D, Lewis J, Pourquie O: **Control of segment number in vertebrate embryos.** *Nature* 2008, **454**:335-339.
8. Cohn MJ, Tickle C: **Developmental basis of limblessness and axial patterning in snakes.** *Nature* 1999, **399**:474-479.
9. Kohlsdorf T, Cummings MP, Lynch VJ, Stopper GF, Takahashi K, Wagner GP: **A molecular footprint of limb loss: sequence variation of the autopodial identity gene Hoxa-13.** *J Mol Evol* 2008, **67**:581-593.
10. Vonk FJ, Admiraal JF, Jackson K, Reshef R, de Bakker MA, Vanderschoot K, van den Berge I, van Atten M, Burgerhout E, Beck A, Mirtschin PJ, Kochva E, Witte F, Fry BG, Woods AE, Richardson MK: **Evolutionary origin and development of snake fangs.** *Nature* 2008, **454**:630-633.
11. Di-Poi N, Montoya-Burgos JI, Miller H, Pourquie O, Milinkovitch MC, Duboule D: **Changes in Hox genes' structure and function during the evolution of the squamate body plan.** *Nature* 2010, **464**:99-103.
12. Moon BR, Tullis A: **The ontogeny of contractile performance and metabolic capacity in a high-frequency muscle.** *Physiol Biochem Zool* 2006, **79**:20-30.
13. Matsubara K, Tarui H, Toriba M, Yamada K, Nishida-Umehara C, Agata K, Matsuda Y: **Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes.** *Proc Natl Acad Sci U S A* 2006, **103**:18190-18195.
14. Sparkman AM, Arnold SJ, Bronikowski AM: **An empirical test of evolutionary theories for reproductive senescence and reproductive effort in the garter snake *Thamnophis elegans*.** *Proc Roy Soc B Biol Sci* 2007, **274**:943-950.
15. Geffeney S, Ruben PC, Brodie ED Jr, Brodie ED III: **Mechanisms of adaptation in a predator-prey arms race: TTX resistant sodium channels.** *Science* 2002, **297**:1336-1339.
16. Castoe TA, de Koning APJ, Kim HM, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD: **Evidence for an ancient adaptive episode of convergent molecular evolution.** *Proc Natl Acad Sci U S A* 2009, **106**:8986-8991.
17. Castoe TA, Jiang ZJ, Gu W, Wang ZO, Pollock DD: **Adaptive evolution and functional redesign of core metabolic proteins in snakes.** *PLoS ONE* 2008, **3**:e2201.
18. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E: **Early evolution of the venom system in lizards and snakes.** *Nature* 2006, **439**:584-588.
19. Cruz LS, Vargas R, Lopes AA: **Snakebite envenomation and death in the developing world.** *Ethn Dis* 2009, **19(1 Suppl 1)**:S1-42-46.
20. Castoe TA, Gu W, de Koning AP, Daza JM, Jiang ZJ, Parkinson CL, Pollock DD: **Dynamic nucleotide mutation gradients and control region usage in squamate reptile mitochondrial genomes.** *Cytogenet Genome Res* 2009, **127**:112-127.
21. **Snake Genomics** [http://www.snakegenomics.org]
22. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, Harris JR, Ehrlich SD, Apweiler R, Austin CP, Berglund L, Bobrow M, Bountra C, Brookes AJ, Cambon-Thomsen A, Carter NP, Chisholm RL, Contreras JL, Cooke RM, Crosby WL, Dewar K, Durbin R, Dyke SO, Ecker JR, El Emam K, Feuk L, Gabriel SB, Gallacher J, Gelbart WM, *et al.*: **Prepublication data sharing.** *Nature* 2009, **461**:168-170.
23. Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD: **Eukaryotic genome size databases.** *Nucleic Acids Res* 2007, **35**:D332-D338.
24. Singh L, Sharma T, Ray-Chaudhuri SP: **Chromosomes and the classification of the snakes of the family Boidae.** *Cytogenetics* 1968, **7**:161-168.
25. Olmo E, Stingo V, Odierna G, Cobror O: **Variations in the repetitive DNA and evolution in reptiles.** *Comp Biochem Physiol* 1981, **69**:687-691.
26. Olmo E: **Genomic composition of reptiles: evolutionary perspectives.** *J Herpetol* 1984, **18**:20-32.
27. Castoe TA, Hall K, Guibotsy Mboulas ML, Gu W, de Koning AP, Poole AW, Vemulapalli V, Daza JM, Feschotte C, Pollock DD: **Discovery of highly divergent repeat landscapes in snake genomes using high throughput sequencing.** *Genome Biol Evol* 2011, doi:10.1093/gbe/evr043.
28. Shedlock AM, Botka CW, Zhao S, Shetty J, Zhang T, Liu JS, Deschavanne PJ, Edwards SV: **Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome.** *Proc Natl Acad Sci U S A* 2007, **104**:2767-2772.
29. Castoe TA, Bronikowski AM, Brodie ED 3rd, Edwards SV, Pfrender ME, Shapiro MD, Pollock DD, Warren WC: **A proposal to sequence the genome of a garter snake (*Thamnophis sirtalis*).** *Stand Genomic Sci* 2011, **4**:257-270.
30. **Amplicon Express** [http://www.genomex.com/]

31. **Ensembl** [http://www.ensembl.org]
32. **UCSC genome browser** [http://genome.ucsc.edu]
33. **NCBI Genomes & Maps** [http://www.ncbi.nlm.nih.gov/guide/genomes]
34. Genome 10K Community of Scientists: **Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species.** *J Hered* 2009, **100:**659-674.