Improved draft of the Mojave Desert tortoise genome, *Gopherus agassizii*, version 1.1

Timothy H. Webster[1†*], Greer A. Dolby[1†*], Melissa A. Wilson Sayres[1,2], Kenro Kusumi[1]

[1]School of Life Sciences, Arizona State University, Tempe, AZ 85287

[2]Center for Evolution and Medicine, Arizona State University, Tempe, AZ 85287

[†]*authors contributed equally*

Correspondence to (*):

Timothy Webster
School of Life Sciences
Arizona State University
PO Box 874501
Tempe, AZ 85287
Timothy.H.Webster@asu.edu

Greer Dolby
School of Life Sciences
Arizona State University
PO Box 874501
Tempe, AZ 85287
gadolby@asu.edu

Running title: Mojave Desert tortoise genome v1.1

Key words: genome, assembly, tortoise, scaffold, contamination

ABSTRACT

Exogenous sequence contamination presents a challenge in first-draft genomes because it can lead to non-contiguous, chimeric assembled sequences. This can mislead downstream analyses reliant on synteny, such as linkage-based analyses. Recently, the Mojave Desert Tortoise (*Gopherus agassizii*) draft genome was published as a resource to advance conservation efforts for the threatened species and discover more about chelonian biology and evolution. Here, we illustrate steps taken to improve the desert tortoise draft genome by removing contaminating sequences—actions that are typically carried out after the initial release of a draft genome assembly. We used information from NCBI's Vecscreen output to remove intra-scaffold contamination and trim heading and trailing Ns. We then reordered and renamed scaffolds, and transferred the gene annotation onto this assembly. Finally, we describe the tools developed for this pipeline, freely available on Github (https://github.com/thw17/G_agassizii_reference_update), which facilitate post-assembly processing of other draft genomes. The new gopAga1.1 genome has an N50 of 251 kb, L50 of 2592 scaffolds, and its annotation retains 17,201 of the original 20,172 genes that were unaffected by the scaffold processing.

INTRODUCTION

The Mojave Desert Tortoise, *Gopherus agassizii*, is a long-lived, xeric-adapted species endemic to southern California, southern Nevada, southwestern Utah, and northwestern Arizona (Morafka & Berry 2002; Murphy *et al.* 2011). One of six extant species in the genus *Gopherus*, it is thought to have diverged from the lineage leading to *G. evgoodei* and *G. morafkai* between 5–6 million years ago when the Colorado River first began draining into the Gulf of California

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

(Dorsey *et al.* 2011; Murphy *et al.* 2011; Edwards *et al.* 2016). These three species have since differentially adapted to their respective habitats, with the differences between *G. agassizii* of the Mojave Desert and *G. morafkai* of the Sonoran Desert being well-characterized (Edwards *et al.* 2015). Differences between these deserts based on seasonal rainfall, total annual precipitation, vegetation, and other key environmental characteristics likely underlie the differential adaptations in these species (Pianka 1970; Reynolds *et al.* 2004) .

Significant conservation efforts have targeted *Gopherus agassizii* since its Threatened listing under the Endangered Species Act in 1990 (Smith 1990). However, populations continue to decline due to a combination of habitat loss, changes in land use, invasive grasses (Drake *et al.* 2016), and upper respiratory tract disease (URTD; (Jacobson *et al.* 1991; Doak *et al.* 1994; Brown *et al.* 1994). As part of this conservation effort, Tollis *et al.* (2017) published a draft genome (version 1.0; gopAga1) of *G. agassizii,* which was the first for any tortoise species. Analysis of the genome revealed putative genes under selection in *G. agassizii* relative to other non-avian reptiles, confirmed slow mutation rates among chelonians (Shaffer *et al.* 2013), and found evidence of gene structure more closely resembling chicken than other non-avian reptiles (Tollis *et al.* 2017).

Development of the reference genome for this species enables new and promising avenues of research that will aid its conservation. Here, we present genome version 1.1 for *G. agassizii* (gopAga1.1), with the following improvements from initial release (gopAga1): **1)** screening for and removal of exogenous contaminant sequences; **2)** reordering and renaming of scaffolds within the assembly; and **3)** an updated annotation that converts the physical positioning of genes and gene features under this new scaffolding. Draft genomes of non-model organisms are rapidly becoming more common and they represent the foundation for future

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

research. Because many such assemblies contain contamination (Alkan *et al.* 2010), software tools and workflows designed to handle the splitting, sorting, and processing of scaffolds are needed. In addition to introducing genome version 1.1 for *G. agassizii*, we provide software tools to manipulate early-generation genome assemblies such as this one, and aim to add transparency to the steps involved in processing a draft assembly to meet the standards required for deposition in public databases (e.g., NCBI).

MATERIALS & METHODS

After submitting gopAga1 for processing and hosting, NCBI (https://www.ncbi.nlm.nih.gov) identified adapter and exogenous sequence contamination using their Vecscreen (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) pipeline, an issue common to many draft genome assemblies. As a part of their pipeline, NCBI removed contaminant sequences from the beginning and ends of scaffolds and provided the locations of remaining contaminants. We used the scripts presented in this manuscript, the processed assembly file, and contamination file to: **1)** split scaffolds at the intra-scaffold sites of contamination provided in the Vecscreen output; **2)** soft-clip scaffold ends that contained Ns after splitting; **3)** reorder and rename v1.1 scaffolds by descending size; **4)** transfer the v1 annotation to v1.1 assembly under these newly processed scaffolds (Figure 1).

Genome assembly version 1.1

Within-scaffold regions of contamination likely resulted in misjoining non-contiguous regions. To remove such effects, we wrote a Python script (Remove_and_split_contamination_NCBI.py) to read NCBI output (with 1-based coordinates)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

and identify contaminated regions in the assembly. We used the script to remove these contaminant sequences and split scaffolds at locations of contamination. For example, a 100-base scaffold with contamination from 15 through 30 would be split into two scaffolds—one 14 bases long (corresponding to bases 1–14) and one 70 bases long (corresponding to bases 31–100; Figure 2). We ran this script with the following command line:

python Remove_and_split_contamination_NCBI.py --fasta GopAga1.0_NCBIout.fasta --output GopAga1.1_nocontam.fasta --ncbi_tab RemainingContamination.txt --wrap_length 90 --delimiters ".." "," --fasta_id_junk "lcl|"

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

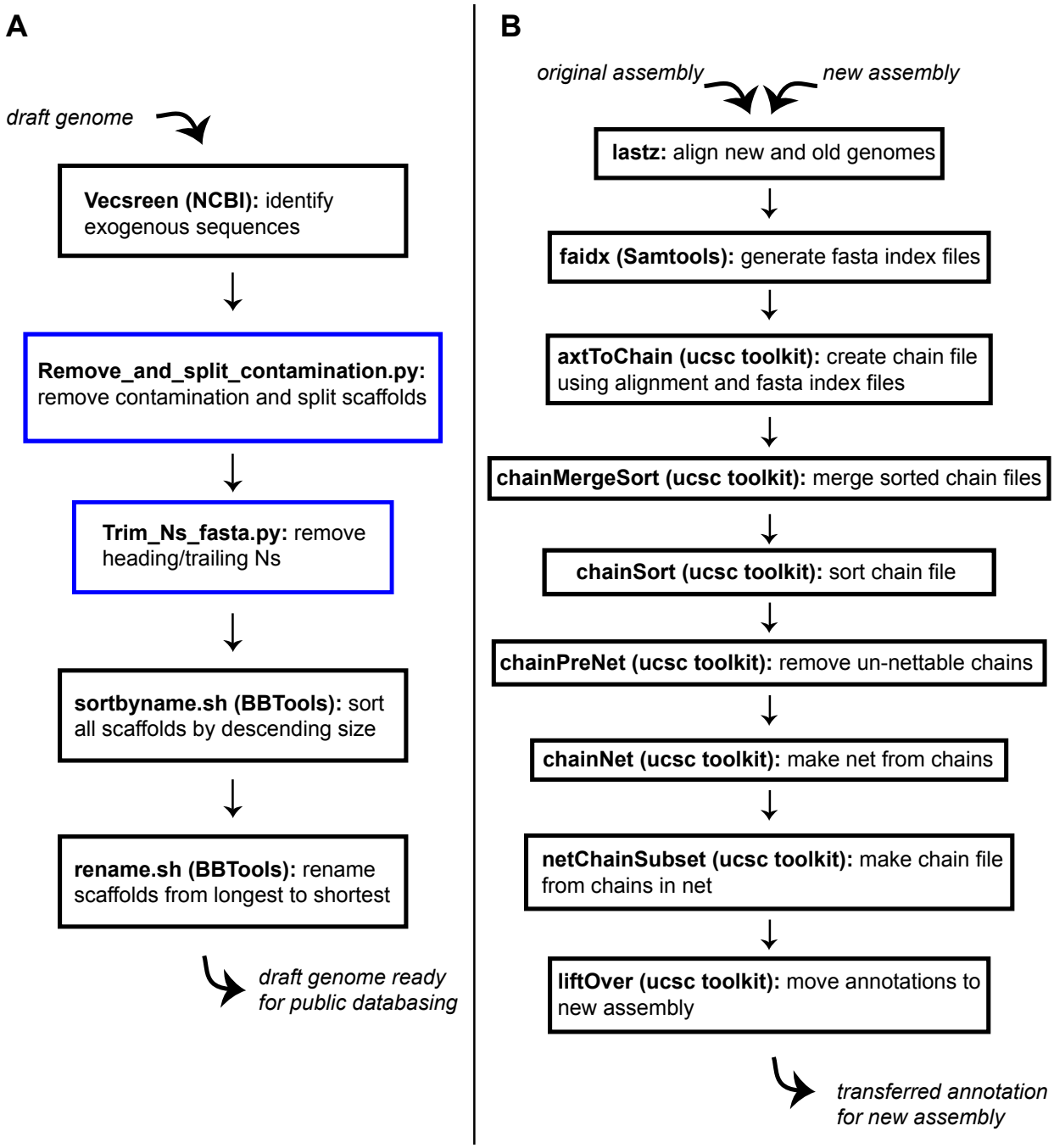

Figure 1. Overview of assembly (A) and annotation (B) processes used for gopAga1.1. Blue boxes are scripts presented here; black boxes are tools provided by other software packages. Descriptions of external tools are reproduced here from their original source documentation.

108

109

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

When the estimated physical distance between regions is known (e.g., through mate-pair sequencing), Ns are often used to fill in unknown sequences between contigs. Many contamination sites were adjacent to these strings of Ns, suggesting at least some contamination was introduced during scaffolding steps. In these cases, splitting scaffolds at the sites of contamination left the newly split scaffolds with long strings of either leading or trailing strings of Ns. Using a second Python script (Trim_Ns_fasta.py), we dynamically trimmed these patterns and removed any remaining contigs and scaffolds less than 100 bp in length with the following command line:

python Trim_Ns_fasta.py --fasta INFILE.fasta --output_fasta OUTFILE.fasta --filtered_scaffolds removed_scaffolds.txt --wrap_length 60 --soft_buffer 10 --min_n 1 --minimum_length 100

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

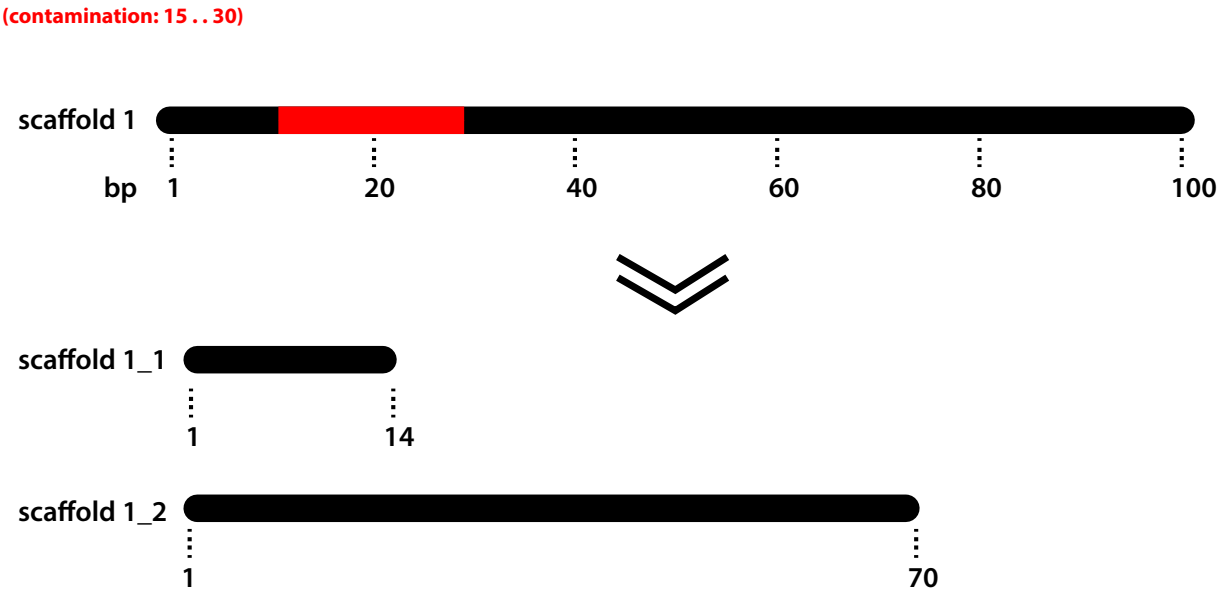

Figure 2. Schematic showing how the contaminant removal and scaffold splitting processes work. The contaminated regions (red) provided by Vecscreen output are inclusive, 1-based coordinates and are removed, leaving two new unassociated scaffolds. New scaffolds are named numerically using the original scaffold number (e.g., scaffold 1_1 and scaffold 1_2).

The *soft_buffer* parameter directs the program to remove up to and including the provided length before looking for Ns. However, no clipping will occur if an N is not discovered. A --*hard_buffer* option is also implemented in the program, which will instead hard-clip a sequence by a certain length, whether or not an N is discovered.

We then used *sortbyname.sh*, in the BBTools suite (https://sourceforge.net/projects/bbmap), to sort scaffolds by descending length with the command:

sortbyname.sh in=GopAga1.1_unsorted.fasta out=GopAga1.1.sorted.fasta length descending

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

We used *rename.sh*, also part of the BBTools suite (https://sourceforge.net/projects/bbmap), to rename the split and sorted scaffolds for version 1.1. When draft genomes are assembled using multiple software tools, it can result in subtly different scaffold naming schemes that can cause confusion (i.e., scaffold_412 vs. scaffold412 vs. Scaffold412). Here we used increasing numbers for scaffold names corresponding to decreasing length. As such, the longest scaffold is named scaffold_0, the next longest is scaffold_1, and so on. We achieved this using the following command:

rename.sh in=GopAga1.1.sorted.fasta out=GopAga1.1.sorted.renamed.fasta prefix=scaffold

We performed manual quality control assessments at each step in these processes. Such assessments included visual examination of split and excised scaffold regions in comparison to Vecscreen output, comparing number of scaffolds pre- and post-splitting to the number of contaminated regions, visual examination of pre- and post-soft-clipped scaffolds, comparing file sizes before and after each step, and comparing checksums when moving files. We also used standard UNIX tools to count scaffolds, changes in nucleotide composition, and check scaffold names. Finally, we compared sequence statistics between the two assemblies using *stats.sh* in the BBTools suite (https://sourceforge.net/projects/bbmap). We only modified the "minscaf" parameter to calculate statistics with different minimum scaffold sizes. Note that in this manuscript, we use "scaffold" in a broad sense, to refer to any sequence in the assembly with an identifier. This will include both scaffolds containing contigs joined during a scaffolding process and unscaffolded contigs.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Annotation version 1.1

The annotation for version 1 was generated using *ab initio* gene model predictions combined with deep transcriptome mRNA transcription evidence from four adult tissues, including blood, brain, lung, and skeletal muscle (Tollis *et al.* 2017). This original annotation produced a similar number of protein-coding genes (20,172) to western painted turtle and Chinese softshell turtle (21,796 and 19,327, respectively; (Shaffer *et al.* 2013; Wang *et al.* 2013). As part of gopAga1.1 we lifted this *de novo* annotation for gopAga1 onto the gopAga1.1 assembly using the following methodology. First, we aligned genome assemblies for versions 1.0 and 1.1 using *lastz_32* (Harris 2007). After trying several different combinations of parameters, the best results were produced using gapping, nochain, nogfextend, mismatch=(0,100), exact=20, step=30, notransition, notwins, traceback=160.0M and seed=match12 with output format as .axt. Importantly, the v1.1 assembly is a subset of v1.0 and has no nucleotide differences aside from the removal of contamination and training Ns, which may be an uncommon scenario for these alignment tools. We converted the output alignment (.axt) file to a chain file using the *axtToChain* tool from ucsc toolkit (http://genome.cse.ucsc.edu/index.html). We sorted the chain file by score using *chainSort* and removed chains that would not be netted using *chainPreNet.* We performed netting with *netChainSubset* to create larger blocks of chains and used the –skipMissing parameter because our chains were filtered. Using this final chain file, we lifted over the annotation using *liftOver* from ucsc toolkit.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

Removing small scaffolds

NCBI submission requires assemblies to contain only scaffolds with sequence lengths greater than or equal to 200 nucleotides. To filter the FASTA assembly itself, we used bioawk (https://github.com/lh3/bioawk):

bioawk -c fastx '(length($seq) > 199) {print ">"$name"\n"$seq }' GopAga1.1.sorted.renamed.fasta > GopAga1.1.sorted.renamed.min200.fasta

We then created a BED file of scaffolds removed in the above command, determined by comparing fasta indexes generated with SAMtools faidx (Li *et al.,* 2009), and used BEDTools (Quinlan & Hall 2010) to subtract annotations on these filtered scaffolds:

bedtools subtract –A –a GopAga1.1.annotation_final.gff –b GopAga1.1_min200.bed > GopAga1.1.annotation_final_above200.gff

Data and Software availability

The Python scripts described above, *Remove_and_split_contamination.py* and *Trim_Ns_fasta.py*, are freely available on Github (https://github.com/thw17/G_agassizii_reference_update) and in the Supporting Information. We have deposited the fasta sequence and annotation files for gopAga1.1 in the Harvard Dataverse (doi:10.7910/DVN/HUASUW). The fasta sequence is available on NCBI under BioProject PRJNA352726, BioSample SAMN05991319, and accession PPEB00000000.1 (scaffold accessions PPEB01000001-PPEB01172559).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

RESULTS & DISCUSSION

GopAga1 and gopAga1.1—which we alternatively refer to in this manuscript as v1.0 and v1.1, respectively—differ in a few important ways. First, we removed contaminant sequences (primarily adapters) present in v1.0 and split scaffolds around sites of contamination. We removed leading and trailing Ns from sequences before sorting and renaming scaffolds by size. Finally, we removed all sequences smaller than 200 bases and lifted over the annotation to the modified assembly. We outline the differences in resulting sequence statistics between v1.0 and v1.1 below.

Genome assembly version 1.1

While splitting scaffolds at sites of contamination initially increased the number of scaffolds, removing scaffolds less than 200 bases led to a major overall reduction in the number of scaffolds (v1.0: 863,216; v1.1: 172,559; Table 1). These procedures also led to a reduction in assembly size, from 2.399 Gb in v1.0 to 2.184 Gb in v1.1 (Table 1). Of the removed sequences, approximately 58% consisted of either contamination or Ns, while the remaining 42% were removed because they were under the 200 bp threshold.

Filtering and trimming also affected other genome statistics. We measured N50 (more than 50% of the genome is found in scaffolds this size scaffold or larger) and L50 (minimum number of scaffolds containing 50% or more of the total sequence length) on scaffolds greater than 200 bp in versions 1.0 and 1.1. In v1.0, the N50 was 251 kb and L50 was 2592 scaffolds (Table 1); in v1.1, N50 was 228 kb and L50 was 2740 scaffolds (Table 1). This effect was largely driven by splitting some larger scaffolds that may have been joined by contamination because the differences between v1.0 and v1.1 are more pronounced when only considering

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

scaffolds longer than 200 bp, which were the scaffolds primarily affected in the contaminant processing (v1.0: N50 = 265 kb, L50 = 2418; v1.1: N50 = 228 kb, L50 = 2740; Table 1).

Table 1. Comparison of draft assembly statistics.

	gopAga1.0_min0[a]	***gopAga1.0_min200***[b]	***gopAga1.1***[c]
Total Length[d]	2,399,952,228	2,309,856,185	2,184,968,471
Num. Scaffolds[e]	863,216	189,565	172,559
Longest Scaffold[f]	2,046,553	2,046,553	1,743,037
L50/N50[g]	2592/251 kb	2418/265 kb	2740/228 kb
L90/N90[h]	13,331/19 kb	10,799/35 kb	10647/43 kb
%GC[i]	43.85%	43.70%	43.62%
%N[j]	1.55%	1.61%	1.51%

[a]Version 1.0 of the *G. agassizii* genome containing all scaffolds.
[b]Version 1.0 of the *G. agassizii* genome containing only scaffolds greater than 199 bp.
[c]Version 1.1 of the *G.agassizii* genome (which contains only scaffolds greater than 199 bp).
[d]Total length of the assembly, including Ns.
[e]Total number of named scaffolds in the assembly.
[f]Sequence length of the longest scaffold in the assembly.
[g]L50 is the minimum number of scaffolds containing 50% or more of the assembly. 50% of the genome is found in scaffolds of length N50 or greater.
[h]L90 is the minimum number of scaffolds containing 90% or more of the assembly. 90% of the genome is found in scaffolds of length N90 or greater.
[i]Percent of total sequence that is G or C.
[j]Percent of total sequence that is N.

Annotation version 1.1

Annotation of the draft genome v1.0 identified 20,172 genes, of which 17,201 are present in the v1.1 assembly. Of the 2,971 genes not lifted over in the v1.1 genome, 2,731 of those failed to lift over due to being split across scaffolds in the v1.1 assembly, 118 were partially deleted in the v1.1 assembly, and 122 were fully deleted in the new scaffolds. These results indicate that

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

the assembly and/or annotation of some genic or gene-associated regions may have been influenced by exogenous sequence in the v1.0 assembly, and may reflect a common challenge of draft genomes.

CONCLUSIONS

The draft genome of *Gopherus agassizii,* the first tortoise sequenced, advances conservation biology and management of this species and comparative genomic studies. Here we improve the *G. agassizii* draft genome in version 1.1. Improvements include removing contamination, splitting scaffolds at sites of internal contamination, reordering and renaming scaffolds, and transferring the v1.0 annotation coordinates to v1.1. We include scripts and detailed commands to aid in processing other draft genomes, which often require similar filtering and restructuring, particularly for deposition into public databases. A particularly important message is that adaptor contamination can present a major challenge for short read assemblers, causing reads and contigs to misassemble (Alkan *et al.* 2010; Schmieder & Edwards 2011; Bolger *et al.* 2014) and leading to errors in contiguity and/or synteny. Generally speaking, these errors fell in intergenic regions, though a number of genes were impacted. Care must be taken to include an exhaustive list of adaptors used by sequencing projects to ensure that trimming programs are using all potentially relevant sequences.

We believe that the continued development of this resource will enable new, promising directions in tortoise research and conservation. In particular, this resource allows reconstructions of modern and historical demographic patterns with greater statistical power. It enables researchers to disentangle the history of gene flow and ecological adaptations that differentiate *G. agassizii* from *G. morafkai*. And finally, it can aid in the characterization of the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

immune system of chelonians, leading to a better understanding of why URTD affects tortoise species differently and development of better diagnostics for detection, which would benefit management of the species.

ACKNOWLEDGEMENTS

We thank Marc Tollis, Dale DeNardo, John Cornelius, Taylor Edwards, Cristina Jones, and Mariana Grizante Bortoletto for helpful conversations and ongoing collaborations.

AUTHOR NOTE

This manuscript is intended to detail the changes made between version 1.0 (Tollis *et al*., 2017) and 1.1 of the Mojave Desert tortoise genome (available on NCBI). This manuscript will not be submitted for peer review, but genome version 2.0 is forthcoming.

REFERENCES

Alkan C, Sajjadian S, Eichler EE (2010) Limitations of next-generation genome sequence assembly. *Nature Methods* 8:61–65.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Brown MB, Schumacher IM, Klein PA *et al.* (1994) *Mycoplasma agassizii* causes upper respiratory tract disease in the desert tortoise. *Infection and Immunity* 62:4580–4586.

Doak D, Kareiva P, Klepetka B (1994) Modeling population viability for the Desert Tortoise in the western Mojave Desert. *Ecological Applications* 4:446–460.

Dorsey RJ, Housen BA, Janecke SU, Fanning CM, Spears ALF (2011) Stratigraphic record of basin development within the San Andreas fault system: Late Cenozoic Fish Creek–Vallecito basin, southern California. *Geological Society of America Bulletin* 123:771–793.

Drake KK, Bowen L, Nussear KE *et al.* (2016) Negative impacts of invasive plants on conservation of sensitive desert wildlife. *Ecosphere* 7:e01531–20.

Edwards T, Berry KH, Inman RD *et al.* (2015) Testing taxon tenacity of tortoises: evidence for a geographical selection gradient at a secondary contact zone. *Ecology and Evolution* 5:2095–2114.

Edwards T, Tollis M, Hsieh P *et al.* (2016) Assessing models of speciation under different biogeographic scenarios; an empirical study using multi-locus and RNA-seq analyses.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

306 *Ecology and Evolution* 102:1–18.
307 Harris RS (2007) Improved pairwise alignment of genomic DNA. D. Phil. Thesis. The
308 Pennsylvania State University.
309 Jacobson ER, Gaskin JM, Brown MB *et al.* (1991) Chronic Upper Respiratory Tract Disease of
310 free-ranging desert tortoises (*Xerobates agassizii*). *Journal of Wildlife Diseases* 27:296–316.
311 Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and
312 SAMtools. *Bioinformatics* 25:2078–2079.
313 Morafka DJ, Berry KH (2002) Is *Gopherus agassizii* a desert-adapted tortoise or an exaptive
314 opportunist? Implications for tortoise conservation. *Chelonian Conservation and Biology*
315 4:263–287.
316 Murphy R, Berry K, Edwards T *et al.* (2011) The dazed and confused identity of Agassiz's land
317 tortoise, *Gopherus agassizii* (Testudines: Testudinidae) with the description of a new species
318 and its consequences for conservation. *Zookeys* 113:39–71.
319 Pianka ER (1970) Comparative autecology of the lizard *Cnemidophorus tigris* in different parts
320 of its georgraphic range. *Ecology* 51:703–720.
321 Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic
322 features. *Bioinformatics* 26:841–842.
323 Reynolds JF, Kemp PR, Ogle K, Fernández RJ (2004) Modifying the "Pulse-Reserve" paradigm
324 for deserts of North America: precipitation pulses, soil water, and plant responses. *Oecologia*
325 141:194–210.
326 Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from
327 genomic and metagenomic datasets (F Rodriguez-Valera, Ed,). *PLoS ONE* 6:e17288–11.
328 Shaffer HB, Minx P, Warren DE *et al.* (2013) The western painted turtle genome, a model for
329 the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome*
330 *Biology* 14:R28.
331 Smith R (1990) Endangered and threatened wildlife and plants; determination of threatened
332 status for the Mojave population of the desert tortoise. *Federal Registrar* 55:12178–12191.
333 Tollis M, DeNardo DF, Cornelius JA *et al.* (2017) The Agassiz's desert tortoise genome provides
334 a resource for the conservation of a threatened species. *PLoS ONE* 12:e0177708.
335 Wang Z, Pascual-Anaya J, Zadissa A *et al.* (2013) The draft genomes of soft-shell turtle and
336 green sea turtle yield insights into the development and evolution of the turtle-specific body
337 plan. *Nature Genetics* 45:701–706.
338
339

340 **AUTHOR CONTRIBUTIONS**

341 THW and GAD conducted analyses. GAD and THW wrote the manuscript. KK and MWS
342 provided oversight and computing resources. THW, GAD, MWS, and KK edited the manuscript.

343

344

345

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3266v4 | CC BY 4.0 Open Access | rec: 31 Jan 2018, publ: 31 Jan 2018

FUNDING

GAD and KK were supported by a US Geological Survey Cooperative Ecosystem Studies Units (CESU) award, GAD and THW were supported by a Fostering Postdoctoral Research in the Life Sciences seed grant from the School of Life Sciences at Arizona State University. All authors were supported by a Heritage Grant from the Arizona Game and Fish Department. This work was supported by funding from the College of Liberal Arts and Sciences at ASU to KK. Computational analysis was supported by allocations from Research Computing at Arizona State University.