bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The landscape of structural variation in coppery titi monkeys (*Plecturocebus cupreus*)

Cyril J. Versoza[1], Karen L. Bales[2–4], Jeffrey D. Jensen[1], Susanne P. Pfeifer[1,*]

[1] Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, USA

[2] Department of Psychology, University of California, Davis, CA, USA

[3] California National Primate Research Center, Neuroscience and Behavior Division, Davis, CA, USA

[4] Department of Neurobiology, Physiology, and Behavior, University of California, Davis, CA, USA

[*] corresponding author: susanne@spfeiferlab.org

keywords: primate; platyrrhine; Pitheciidae; copy number variation; structural variation; population genomics

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Coppery titi monkeys (*Plecturocebus cupreus*) are an important non-human primate model for studying neurobiology and social behavior, in part owing to their relatively unusual combination of social monogamy and paternal care. Despite this importance, relatively little is known regarding the underlying population genomics of this platyrrhine. This study presents high-coverage, whole-genome sequencing data from 26 individuals which, combined with a highly accurate multi-algorithm ensemble approach, was used to characterize the first map of structural variation in the species. This novel genomic resource includes over 13,000 structural variants, with the majority (>90%) being copy number variants. While many of these were found to be located in intergenic regions, several affected genes associated with disease, including an inversion predicted to impact a pathway implicated in early-onset Parkinson's disease. Furthermore, utilizing parent-offspring trios included within this study, the *de novo* structural variant rate was estimated to be one in every 1.5 births, similar to that reported in rhesus macaques but considerably higher than that observed in large human cohorts, as may be expected from underlying differences in life history traits amongst these species. Taken together, these insights into the structural variant landscape of *P. cupreus* will not only improve their utility as a behavioral model system, but will also contribute to our general understanding of the role of structural variation in both the evolution of the primate clade and disease-outcomes.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Copy number variants — genomic segments larger than 50 bp that vary in DNA dosage between individuals (including deletions and duplications) — and other structural variants that change genome organization (including insertions and inversions), are one of the largest sources of heritable variation (Conrad et al. 2010; 1000 Genomes Project Consortium 2015; Chaisson et al. 2019; Collins et al. 2020). As such, they play an important role in shaping population genetic and phenotypic variation within and between species (Hollox et al. 2022). In addition to potentially facilitating adaptation in novel and changing environments, the divergence of structural variants amongst populations can also impact the degree of gene flow, and thereby contribute to reproductive isolation and speciation (see the reviews of Iskow et al. 2012 and Kondrashov 2012, and the commentary by Feulner and De-Kayne 2017). Moreover, structural variants impacting coding and regulatory regions can have profound effects on health and disease, particularly with regards to the susceptibility to infectious agents (see the review of Hollox and Hoh 2014) as well as neurodevelopmental, neurological, and psychiatric disorders including autism spectrum disorder, schizophrenia, attention deficit hyperactivity disorder, Alzheimer's and Parkinson's disease (see the review of Harner et al. 2025 and references therein).

Historically studied using molecular cytogenic techniques (such as fluorescence in situ hybridization and comparative genomic hybridization) and chromosomal microarray technologies, the emergence of whole-genome sequencing technologies has recently provided a valuable and cost-effective alternative to identify copy number and other structural variants at the population-scale. For paired-end short-read data specifically — which remains the *de facto* standard in the field due to both lower costs and sample requirements relative to those necessary for long-read technologies — several structural variant callers have been developed that scan the genome for regions that exhibit variations in read coverage, harbor an excess of split read alignments, and/or read pairs with discordant insert sizes or conflicting strand orientation (for an overview of the different detection strategies, see Figure 2 in Alkan et al. 2011). In contrast to the more widely

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

studied single nucleotide variation (Pfeifer 2017), standardized guidelines still need to be established for the identification of structural variation (Ho et al. 2020); however, while individual short-read structural variant callers can suffer from high false discovery rates, the application of multiple methodologies leveraging different (and often complementary) signals — referred to as an "ensemble" approach — has generally been shown to improve both sensitivity and specificity compared to single-caller approaches (see the benchmarking studies of Cameron et al. 2017; Kosugi et al. 2019; Gabrielaite et al. 2021; Kosugi and Terao 2024). Indeed, recent research has demonstrated that, while long-read data is necessary to comprehensively characterize structural variants in repetitive regions of the genome, short-read data exhibit a similar level of precision and recall for different types of structural variants in non-repetitive regions as well as for large (> 1 kb) deletions in repetitive regions (see Figure 4 in Kosugi and Terao 2024).

Due to these computational and technological advances, recent years have witnessed a renewed interest in the study of structural variation in many organisms. In primates, for example, differences in copy number variation have been observed between the genomes of anthropoid apes and other haplorrhines, contributing to the divergence between species (e.g., Gokcumen et al. 2013; Sudmant et al. 2013; Dennis et al. 2017; Kronenberg et al. 2018; Li et al. 2020b). In two of the most comprehensive studies of structural variation in primates to date, Porubsky et al. (2020) used single-cell and long-read sequencing to catalogue inversions across the great apes, whereas Mao et al. (2024) combined information from genome assemblies of eight non-human primates (chimpanzee, bonobo, gorilla, orangutan, gibbon, macaque, owl monkey, and marmoset) together with short- and long-read sequencing data to catalogue deletions and insertions fixed in the human lineage, showing that >25% of the genome has been impacted by structural variation. While long-read sequencing data remains limited for many species, population-specific catalogs of structural variation have also been generated from short-read data for several organisms. For instance, Brasó-Vives et al. (2020) and Thomas et al. (2021) constructed genome-wide maps of copy number variation from 198 and 32 rhesus macaques —

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the most frequently used non-human primate model in biomedical research — finding that, similar to humans, a large proportion of the species' genome is impacted by structural variation. Furthermore, comparing the genomes of 60 rhesus macaques from five Chinese populations and one Indian population, Liu et al. (2025) identified several population-specific copy number variants suggested to be related to environmental adaptation and phenotypic variability (such as body size). Similarly, sequencing the genomes of 38 common marmosets — another species of biomedical relevance — Yang et al. (2023) discovered a deletion enriched in individuals displaying an epileptic phenotype. More recently, based on short-read data from 14 individuals, Versoza et al. (2025) provided the first insights into the structural variation characterizing the genome of the aye-aye (and see the review of Soni et al. 2025) — representing one of the most basal splits in the primate tree. Yet, despite these advances over the past years, insights into the landscape of structural variation remains lacking in many species of biomedical, behavioral, and evolutionary interest.

Native to the neotropical forests of Brazil and Peru, diurnal coppery titi monkeys (*Plecturocebus cupreus*; formerly classified as *Callicebus cupreus*; Groves 2005) form family groups consisting of a socially monogamous (bonded) pair and their offspring that often span multiple generations (Kinzey 1997). Females typically give birth to a single offspring once per year; similar to other primates, coppery titi monkeys have altricial young, though a defining feature of their social system is the extensive involvement of fathers in infant care (Mendoza and Mason 1986; Valeggia et al. 1999). The combination of social monogamy and paternal care — two biological characteristics uncommon amongst non-human primates and other mammals (Lukas and Clutton-Brock 2013) — make coppery titi monkeys a particularly valuable model system for studying neurobiology and social behavior (Bales et al. 2007, 2021; and see the review of Bales et al. 2017). For example, coppery titi monkeys have been used to study the effects of intranasally administered oxytocin — a neurohormone that plays a crucial role in pair bonding as well as in a variety of other social behaviors (Carter et al. 2020; Bales et al. 2021; Rigney et al. 2022) — and

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

social interactions (Arias-del Razo et al. 2022a,b; Zablocki-Thomas et al. 2023; Witczak et al. 2024), with additional emerging evidence from clinical studies in humans suggesting that intranasally administered oxytocin might be a potential treatment to reduce social impairment in individuals affected by autism spectrum disorder (e.g., Moerkerke et al. 2024; and see the review of Horta et al. 2020). Yet, despite their importance in behavioral research, current knowledge regarding the underlying population genomic variation of coppery titi monkeys remains limited.

Combining novel high-coverage sequencing data from 26 coppery titi monkeys with a previously established, highly accurate multi-algorithm ensemble approach, this study presents the first map of structural variation in the species. Although the majority of identified variants were located in intergenic regions of the genome, several were predicted to be of high importance functionally. Amongst these variants, six impacted genes associated with human disease, including an inversion predicted to affect the PINK1-PARKIN pathway which has previously been linked to early-onset Parkinson's disease. Taken together, these first insights into the structural variant landscape provided here will thus not only improve the usage of coppery titi monkeys as a behavioral model system, but also provide new avenues for future investigations of the biological mechanisms underlying human biology, health and disease.

MATERIALS AND METHODS

Animal subjects

This study was performed in compliance with all regulations regarding the care and use of captive primates, including the NIH Guidelines for the Care and Use of Animals and the American Society of Primatologists' Guidelines for the Ethical Treatment of Nonhuman Primates. Procedures were approved by the UC-Davis Institutional Animal Care and Use Committee (protocol 22523).

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

123

124 **Samples, sequencing, and read mapping**

125 Blood samples of 26 captive coppery titi monkeys (12 females and 14 males) were
126 obtained from the pedigreed colony housed at the California National Primate Research Center.
127 For each sample, genomic DNA was isolated, a PCR-free library was prepared to guard against
128 amplification errors, and the whole genome was sequenced (Illumina NovaSeq 6000; 150 bp ×
129 150 bp) to >50-fold coverage on average. The coverage in this study is thus considerably higher
130 than that generally recommended (20×; Wold et al. 2021), with higher levels of coverage shown
131 to increase both the accuracy and sensitivity of structural variant detection (Ahmad et al. 2023).
132 The resulting sequencing reads were quality-controlled using TrimGalore v.0.6.10
133 (https://github.com/FelixKrueger/TrimGalore) which trims paired-end reads in a synchronized
134 manner, discarding any pairs where at least one of the reads exhibits a length of < 20 bp after
135 trimming Illumina adapter sequences and bases with a quality score < 20 from the 3'-ends. The
136 pre-processed paired-end reads were aligned to the species-specific reference genome
137 (PleCup_hybrid; GenBank accession number: GCA_040437455.1; Pfeifer et al. 2024) using
138 BWA-MEM v.0.7.17 (Li 2013) and the resulting read alignments were sorted by coordinate using
139 SAMtools *sort* v.1.16 (Danecek et al. 2021). Alignments from different runs of the same sample
140 were merged using the Genome Analysis Toolkit (GATK) *MergeSamFiles* v.4.2.6.1 (van der
141 Auwera and O'Connor 2020); subsequently, optical duplicates were flagged using *MarkDuplicates*
142 in order to avoid potential biases in variant calling arising from artificially inflated read coverage
143 (Pfeifer 2017). As high error rates, non-uniform coverage, and irregular insert size distributions
144 can impede the discovery of structural variants (Mahmoud et al. 2019), SAMtools v.1.16 (Danecek
145 et al. 2021) and goleft v.0.2.6 (https://github.com/brentp/goleft) were used to evaluate the quality
146 and coverage patterns of the duplicated-marked mappings, respectively. Due to the higher quality
147 and depth of coverage, analyses presented in this study focused on the autosomes (i.e.,

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

chromosomes 1-22). Sample information and mapping statistics are provided in Supplementary Table 1; and see Supplementary Figure 1 for the read coverage across autosomes.

Structural variant discovery

Structural variants were discovered via an ensemble approach consisting of three short-read callers: DELLY (Rausch et al. 2012), Lumpy (Layer et al. 2014), and Manta (Chen et al. 2016). First, integrating paired-end mapping, split-read, and read depth signals from the read alignments, structural variants were jointly called in the study cohort using DELLY *call* v.1.2.6. Second, relying on the same signals as DELLY, in Lumpy v.0.2.13 (executed via Smoove *call* v.0.2.6; https://github.com/brentp/smoove) breakpoint likelihoods were modelled using a probabilistic framework to jointly call structural variants in the cohort. Based on the evidence observed at these breakpoints, the Bayesian likelihood method SVTyper v.0.7.0 (Chiang et al. 2015) was then used to determine the genotype likelihood for each sample. To limit the number of false positives in the dataset, and following the developers' guidelines: (1) variants with a mean Smoove heterozygote quality (MSHQ) score ≤ 3 were excluded to remove heterozygous sites with a low genotyping confidence, (2) deletions with a < 0.7 fold-change in coverage relative to flanking regions (DHFFC) were filtered out to guard against alignment errors (particularly in regions of low complexity), and (3) duplications with a > 1.3 fold-change in coverage relative to genomic regions with similar GC-content (DHBFC) were removed to guard against artefacts resulting from GC-bias. Third, Manta v.1.6.0 was used to jointly call structural variants in the cohort by initially determining rough candidate breakpoint regions based on paired-end mapping and split-read signals only and subsequently performing a local *de novo* assembly in these candidate regions to resolve the breakpoints at base-pair resolution.

For each tool-specific call set, variants that passed the internal quality criteria were filtered following Wold et al. 2025, excluding any deletions shorter than 50 bp, duplications shorter than 300 bp, and inversions shorter than 300 bp given the difficulty of reliably distinguishing genuine

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

signals from sequencing and alignment errors with short-read data at these scales; additionally, any variants longer than 50 kb were excluded to guard against unresolved repeats and other reference assembly issues. Lastly, tool-specific call sets were merged using SURVIVOR *merge* v.1.0.7 (Jeffares et al. 2017) to obtain consensus calls of structural variants detected by at least two of the three callers (requiring a 500 bp maximum distance between breakpoints of identical type).

Structural variant genotyping

The consensus call set was re-genotyped using Graphtyper2 *genotype_sv* v.2.7.2 (Eggertsson et al. 2019). Following the developers' recommendations, the "aggregate" genotyping model was used for copy number variants and insertions, the "breakpoint" model for inversions, and segregating variants were limited to those for which each individual in the cohort passed all internal sample-level filters. For downstream analyses, the final dataset was separated into sites consistent with the patterns of Mendelian inheritance and those violating these patterns using the *mendelian* plugin in BCFtools v.1.10 (Danecek et al. 2021).

Structural variant annotation

SnpEff v.5.2 (Cingolani et al. 2012) was used to predict the functional impact of the discovered structural variants (note that SnpEff does not provide annotations for insertions). Afterward, structural variants classified as high impact were limited to those exhibiting the highest confidence (following the guidelines for human data; for details see Eggertsson et al. 2019) and then manually curated to evaluate potential medical relevance. This assessment relied on information available in the database of Disease-Gene Associations with annotated Relationships among genes (eDGAR; Babbi et al. 2017), which integrates pathogenicity information from UniProt's Humsavar database (UniProt Consortium et al. 2015), NIH's ClinVar database (Landrum et al. 2016, 2018) and the Online Mendelian Inheritance in Man (OMIM) database

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Amberger et al. 2017). Additionally, a gene enrichment analysis was performed using the DAVID web server (Sherman et al. 2022).

RESULTS AND DISCUSSION

Discovery of structural variants in the coppery titi monkey

Structural variants were discovered by sequencing the genomes of 26 coppery titi monkeys to a mean individual coverage of 51.2× (range: 39.6× to 75.1×; Supplementary Table 1) and applying an ensemble approach that combines three established variant callers (DELLY [Rausch et al. 2012], Lumpy [Layer et al. 2014], and Manta [Chen et al. 2016]) to maximize sensitivity and specificity. This strategy has been demonstrated to perform well in primates (Subramanian et al. 2024; and see the benchmarking studies of Kosugi et al. 2019 and Gabrielaite et al. 2021). To further improve accuracy, discovered variants were re-genotyped using Graphtyper2 (Eggertsson et al. 2019), a population-scale genotyper that has been shown to improve genotype consistency across cohorts using pangenome graphs. Using this approach, a set of 13,492 structural variants were discovered among the 26 individuals analyzed in this study, with a mean of 3,042 variants per individual (range: 2,560–3,340; excluding the individual on which the reference assembly was based upon; Supplementary Table 2), spanning a combined 14.7 Mb of the autosomal genome. This diversity of structural variation is comparable to that observed in rhesus macaques (3,646 structural variants per individual; Thomas et al. 2021) but higher than that observed in humans from short-read data (~2,100 to 2,500 structural variants per individual among the 2,504 individuals from the 26 populations included in the 1000 Genomes Project; Sudmant et al. 2015) as expected given the larger effective population size of coppery titi monkeys compared to humans (Terbot et al. 2026). It should be noted, however, that short-read based studies are unable to capture the full spectrum of structural variation and that more recent

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

studies based on long-read data have been able to identify a much larger number of structural variants in the human genome (~100,000 structural variants among 15 individuals; Audano et al. 2019). Consequently, the number of variants reported here certainly represents an underestimate and future work using cutting-edge long-read technologies will be required to provide a more comprehensive picture of the landscape of structural variation in the species, particularly with regards to large and complex events located within highly repetitive regions for which short-read based approaches have low recall rates (Kosugi and Terao 2024).

Characterization of the landscape of structural variation in the coppery titi monkey

The vast majority of variants were copy number variants (92.8%; 12,484 deletions and 37 duplications), though a small number of unbalanced (765 insertions) and balanced structural variants (206 inversions) were also observed (see Figure 1a for genomic locations). Consistent with previous work showing that short-read approaches are biased towards the discovery of deletions (Pang et al. 2010), more deletions than duplications and insertions were detected in the coppery titi monkey genome. Although the ratio of deletions to duplications and insertions (~16:1) was significantly higher than those observed in humans ($\chi 2$ = 13.303; df = 1; *p*-value = 0.000265) and other haplorrhines (rhesus macaques: $\chi 2$ = 19.438; df = 1; *p*-value = 1.04×10^{-5}; Thomas et al. 2021), it was similar to that previously reported in strepsirrhines (aye-ayes: $\chi 2$ = 1.7751; df = 1; *p*-value = 0.1827; Versoza et al. 2025). In addition to genuine biological variation between the species, differences in both the available species-specific genomic resources and study design likely contribute to the observed differences. First, duplications tend to be poorly resolved in unfinished reference genome assemblies (Hartasánchez et al. 2018); additionally, recent work has shown that duplications and insertions are more difficult to detect than deletions using short-read approaches, with recall rates being inversely correlated to the size of these structural variants (Kosugi and Terao 2024). Consequently, differences in the quality of species-specific reference genomes could potentially lead to an underestimation of duplications in the coppery titi

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

monkey and aye-aye genomes compared to human and rhesus macaque for which complete, or near complete, telomere-to-telomere assemblies are now available (Nurk et al. 2022; Zhang et al. 2025). Second, in contrast to the multi-algorithm ensemble approach applied in the studies of aye-ayes (Versoza et al. 2025) and coppery titi monkeys, the short-read variant catalogs previously generated for humans and rhesus macaques (Thomas et al. 2021) were based on a single caller (Lumpy) which may have missed genuine structural variation present in the genomes of these species (with recent benchmarking demonstrating a mean recall rate of only ~3.5% using Lumpy; see Supplementary Table 10 in Kosugi et al. 2019).

Similar to other species (Thomas et al. 2021; Versoza et al. 2025), the number of structural variants was strongly correlated with chromosomal length (deletions: r = 0.975, p-value = 1.33×10^{-14}; duplications: r = 0.588, p-value = 4.03×10^{-3}; insertions: r = 0.959, p-value = 1.96×10^{-12}; inversions: r = 0.900, p-value = 1.22×10^{-8}; Supplementary Figure 2). Although insertions and inversions were relatively evenly distributed between autosomes (χ^2_{ins} = 13.349, df = 21, p-value = 0.8959 and χ^2_{inv} = 12.032, df = 21, p-value = 0.9388), deletions were significantly enriched on chromosomes 1 (1.17-fold enrichment; FDR-adjusted p-value = 1.16×10^{-6}), 7 (1.13-fold; FDR-adjusted p-value = 8.17×10^{-3}), and 18 (1.24-fold; FDR-adjusted p-value = 1.06×10^{-5}) and depleted on chromosomes 4 (0.87-fold; FDR-adjusted p-value = 6.26×10^{-3}), 14 (0.86-fold; FDR-adjusted p-value = 6.26×10^{-3}), and 17 (0.79-fold; FDR-adjusted p-value = 1.31×10^{-3}) (Figure 1b). Duplications were also significantly enriched on chromosome 1 (7.11-fold; FDR-adjusted p-value = 2.74×10^{-27}); however, this observation should be interpreted with some caution given the small number of duplications in the dataset (n = 37). Moreover, structural variants were not randomly distributed, instead their genome-wide distribution was elevated in telomeric and sub-telomeric regions (Figure 1a) which tend to experience higher rates of non-allelic homologous recombination known to facilitate structural variation, particularly near transposable elements and other repetitive regions (Young et al. 2020).

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Consistent with previous work in primates (Brandler et al. 2016; Brasó-Vives et al. 2020; Thomas et al. 2021; Subramanian et al. 2024; Versoza et al. 2025), both insertions and deletions were relatively short (with median lengths of 61 bp and 309 bp, respectively; Figure 2a) — an observation which likely reflects a combination of both biological constraints and technological limitations. On the one hand, errors during DNA replication and double-strand repair frequently generate short insertions and deletions; additionally, as longer insertions and deletions are more prone to disrupt protein-coding or regulatory regions, cause frameshifts, or change chromatin structure, they are more likely to be deleterious (Taylor et al. 2004; Itsara et al. 2010; Mills et al. 2011; Yang et al. 2024) and are thus expected to be purged from the population via purifying selection. On the other hand, although many short-read structural variant callers perform well in detecting short insertions and deletions, larger events as well as duplications and insertions are inherently more challenging to reliably identify, often leading to an ascertainment bias (Conrad and Hurles 2007; Sudmant et al. 2015; Kosugi et al. 2019; Mahmoud et al. 2019; Delage et al. 2020). In contrast, duplications and inversions typically impacted a greater number of nucleotides per event (with median lengths of 1.2 kb and 2.2 kb, respectively; Figure 2b).

Functional role of structural variants in the coppery titi monkey

In order to predict functional importance, structural variants were intersected with the annotations of the reference genome (Pfeifer et al. 2024) using SnpEff (Cingolani et al. 2012) to identify those impacting protein-coding genes, and the potential medical relevance of large-effect variants was subsequently evaluated using the human database of gene-disease associations, eDGAR (Babbi et al. 2017). The majority of variants was located within intergenic regions (65.08%; Supplementary Table 3), a significant enrichment compared to the overall genome composition ($\chi 2$ = 386,673; df = 6; p-value < 2.2×10^{-16}). Although these variants are expected to be of limited functional importance — and consequently are frequently classified as modifiers or variants of low to moderate effect (Figure 3) — they can nevertheless have profound impacts

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

on gene regulatory landscapes, for example, by altering enhancer / promotor functions or changing the boundaries of topologically associated domains. In fact, recent research has shown that structural variants harbored in intergenic regions frequently experience moderately strong purifying selection, suggesting that they are often detrimental to fitness (Saxena and Baer 2025). Additional modifiers included splice site variants (0.71%) as well as those located within transcripts (0.27%), 3'-UTRs (0.01%) and 5'-UTRs (0.02%). Of the variants residing within genic regions, many were predicted to be of high functional impact. Thereby, structural variants with predicted major effects were significantly enriched in cytoplasmic and membrane-associated cellular components, many of which localized to extracellular exosomes (73.2-fold enrichment; Supplementary Table 4). At the molecular level, they were characterized by enzyme-binding activities (52.7-fold enrichment), with a notable overrepresentation of GTPase activity, GTP-binding, and RNA-binding, consistent with roles in cell signaling, protein transport, and RNA regulation (Wittinghofer and Vetter 2011). Among the structural variants with predicted major effects, six were found to be located within disease-linked genes (Table 1).

A ~2.0 kb inversion was predicted to result in a change of a splice site in PTEN-induced putative kinase 1 (PINK1) which senses mitochondrial damage during cellular stress and that, depending on the severity of the damage, initiates either mitochondrial biogenesis or the removal of dysfunctional mitochondria via mitophagy (O'Callaghan et al. 2023). In humans, mutations in PINK1 have been linked to early-onset Parkinson's disease — a neurodegenerative condition typically characterized by a variety of motor symptoms (such as bradykinesia, rigidity, and tremor) resulting from progressive neuronal cell death (Albanese et al. 2005). Parkinson's disease has an estimated prevalence of 1% in individuals over the age of 65 years (Emborg 2017). Both pathogenicity and zygosity of mutations are known to impact the severity of the disease phenotype; the condition generally manifests around the age of 35 years in homozygous individuals and around the age of 43 years in heterozygous individuals, with females displaying an earlier onset than males (see the review by Yin and Dieriks 2025). Due to their close

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

329 evolutionary relatedness and similar neuroanatomy, behavior, cognitive function, development,
330 motor skills and aging process, non-human primates (including common marmosets, vervet
331 monkeys, cynomolgus and rhesus macaques) are important models to study the environmental
332 and genetic risk factors underlying the disease and to develop neuroprotective and treatment
333 strategies such as dopamine replacement and gene therapies (see the reviews of Emborg 2017;
334 Pan et al. 2024, and references therein). For example, several recent studies used
335 CRISPR/Cas9-mediated knockouts of PINK1 to demonstrate that the disruption of this gene leads
336 to neuronal loss in the cerebral cortex and Parkinson-like symptoms in non-human primates
337 (Yang et al. 2019, 2022; Li et al. 2020a, 2021). Although such studies are frequently based on
338 transgenic or chemically-induced models, recent research suggests that aged non-human
339 primates can also naturally develop symptoms resembling those observed in humans affected by
340 the disease (Hurley et al. 2011; Li et al. 2020a, 2021).

341 A ~7.4 kb inversion was predicted to result in a change of the splice acceptor site in alpha-
342 methylacyl-CoA racemase (AMACR), an enzyme aiding the metabolism of fatty acids and the
343 synthesis of bile acid during digestion. Mutations in this gene have been found to lead to AMACR
344 deficiency in humans which has been linked to a variety of adult-onset neurodegenerative issues
345 ranging from migraines to visual impairments, neuropathy, cognitive decline, and seizures
346 (Ferdinandusse et al. 2000; Thompson et al. 2008; Smith et al. 2010; Dick et al. 2011; and see
347 the review of Clarke et al. 2004). Additionally, there has been evidence that a homozygous
348 mutation in AMACR can cause congenital bile acid synthesis defect type 4, a rare genetic disorder
349 characterized by cholestatic liver disease (Ferdinandusse et al. 2000; Setchell et al. 2003).

350 A 68 bp deletion was predicted to result in the loss of a start codon through a frameshift
351 in the coding region of plakophilin-2 (PKP2) — an integral component of both cardiac
352 desmosomes, cell-cell adhesion complexes that provide both structural and mechanical integrity
353 to the cardiac muscle (Mertens et al. 1996), and the cardiac connexome (Cerrone et al. 2017). In
354 humans, multiple deletions, insertions, non-sense mutations, missense mutations, as well as

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

splice site mutations in this gene are known to cause arrhythmogenic cardiomyopathy — a heart disease with an estimated prevalence of 1 in 1,000 to 5,000 individuals (Basso et al. 2009) that is characterized by defects in cardiac morphogenesis through a fibrofatty replacement of the myocardium, and that is associated with cardiac arrhythmias and sudden cardiac arrest (e.g., Thiene et al. 1988; Gerull et al. 2004; Dalal et al. 2006; van Tintelen et al. 2006; Kirchner et al. 2012; and see the review of Corrado et al. 2017).

A ~0.5 kb deletion was predicted to result in a frameshift in the spalt like transcription factor 4 (SALL4). In humans, mutations in this gene have been implicated in two disorders with similar phenotypes: IVIC syndrome (Paradisi and Arias 2007) and Duane-radial ray syndrome (also known as Okihiro syndrome; Al-Baradie et al. 2002; Kohlhase et al. 2002, 2003; Borozdin et al. 2004; Miertus et al. 2006), both characterized by abnormalities of the upper limbs, motoneuron development as well as a range of other features, often with variable expressivity.

A 1.0 kb deletion was predicted to lead to the loss of an exon in the membrane-anchored GTPase Atlastin-3 (ATL3) gene. Missense mutations in this gene are known to cause a heritable form of sensory neuropathy (hereditary sensory neuropathy type 1F) in humans which can lead to chronic ulcerations, osteomyelitis, and acro-osteolysis of the lower limbs in aging affected individuals (Fischer et al. 2014; Kornak et al. 2014; Xu et al. 2019).

Lastly, a 1.6 kb deletion was predicted to result in a frameshift in the ankyrin repeat domain-containing protein 26 (ANKRD26). In humans, there is evidence that mutations in this gene can cause thrombocytopenia-2 — a condition characterized by an abnormally low number of thrombocytes in the blood of the affected individuals which often causes relatively mild symptoms, including easy bruising and a minor bleeding tendency (Savoia et al. 1999; Drachman et al. 2000).

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

De novo structural variants in the coppery titi monkey

In many species, *de novo* structural variants occur at orders of magnitude lower frequency than point mutations (Belyeu et al. 2021); for example, in humans, high-resolution studies of large cohorts of families observed one *de novo* copy number variant per 3.5 births (with a projected rate of one *de novo* mutation per 2-8 births; Collins et al. 2020). A total of 10 *de novo* structural variants — eight deletions, one duplication, and one insertion — were identified among the parent-offspring trios included in this study, i.e., one in every 1.5 births. This rate is similar to the rate of one in every 1.75 births previously reported in rhesus macaques (i.e., seven deletions and one duplication among 14 trios; Thomas et al. 2021). The higher rate of *de novo* structural mutations observed in coppery titi monkeys compared to those observed in rhesus macaques and humans might, at least in part, be driven by differences in generation time (with species exhibiting shorter generation times expected to experience higher rates of molecular evolution; Ohta 1993). As anticipated from the overall landscape of structural variation, the majority of the *de novo* variants (60%) were located within intergenic regions (Supplementary Table 5); the single exonic variant — a 570 bp deletion — was located within a gene of unknown function (KAL0613339).

CONCLUSION

As a primate of considerable biomedical and behavioral interest, gaining novel insights into the heritable variation characterizing the genome of the coppery titi monkey is crucially important to advance both on-going and future research endeavors related to human health and disease. Moreover, in contrast to haplorrhines, population-scale structural variant catalogs remain scarce in most platyrrhines and thus, this first map of the genomic architecture of structural variation for a representative of the Pitheciidae family will serve as a valuable genomic resource for future evolutionary studies across the primate clade. Nevertheless, even though the high-

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

coverage data presented here offers a first glimpse into the structural variant landscape of the species, future long-read studies will be needed both to characterize the full spectrum of genomic variation — particularly with regards to insertions and inversions as well as large and complex events located within repetitive regions which are systematically under-detected using short-read data — as well as to improve breakpoint resolution, particularly for inversions which are frequently flanked by segmental duplications of near-perfect sequence identity (Porubsky and Eichler 2024). Helpfully, as the cost and sample requirements of long-read sequencing technologies continue to decrease, novel studies are expected to emerge that will allow for a more comprehensive understanding of the causes and consequences of structural variation across different primate lineages.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ACKNOWLEDGEMENTS

DNA extraction, library preparation, and Illumina sequencing were conducted at the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center (supported by NIH Shared Instrumentation Grant 1S10OD010786-01) and Novogene (Sacramento, CA, USA). Computations were performed on the Sol supercomputer at Arizona State University (Jennewein et al. 2023).

FUNDING

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM151008 to SPP and the California National Primate Research Center Pilot Program (NIH P51OD011107). CJV was supported by the National Science Foundation CAREER Award DEB-2045343 to SPP. KLB was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Institute of Mental Health of the National Institutes of Health under Award Numbers R01HD092055 and MH125411, and by the Good Nature Institute. JDJ was supported by National Institutes of Health Award Number R35GM139383. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

CONFLICT OF INTEREST

None declared.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature*. 526(7571): 68–74.

Ahmad SF, Chandrababu Shailaja C, Vaishnav S, Kumar A, Gaur GK, Janga SC et al. (2023) Read-depth based approach on whole genome resequencing data reveals important insights into the copy number variation (CNV) map of major global buffalo breeds. *BMC Genomics*. 24(1): 616.

Al-Baradie R, Yamada K, St Hilaire C, Chan WM, Andrews C, McIntosh N et al. (2002) Duane radial ray syndrome (Okihiro syndrome) maps to 20q13 and results from mutations in SALL4, a new member of the SAL family. *Am J Hum Genet*. 71(5): 1195–1199.

Albanese A, Valente EM, Romito LM, Bellacchio E, Elia AE, Dallapiccola B (2005) The PINK1 phenotype can be indistinguishable from idiopathic Parkinson disease. *Neurology*. 64(11): 1958–1960.

Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet*. 12(5): 363–376.

Amberger JS, Hamosh A (2017) Searching Online Mendelian Inheritance in Man (OMIM): a knowledgebase of human genes and genetic phenotypes. *Curr Protoc Bioinformatics*. 58: 1.2.1–1.2.12.

Arias-del Razo R, Velasco Vazquez ML, Turcanu P, Legrand M, Floch M, Weinstein TAR et al. (2022a) Long term effects of chronic intranasal oxytocin on adult pair bonding behavior and brain glucose uptake in titi monkeys (*Plecturocebus cupreus*). *Horm Behav*. 140: 105126.

Arias-del Razo R, Velasco Vazquez ML, Turcanu P, Legrand M, Lau AR, Weinstein TAR et al. (2022b) Effects of chronic and acute intranasal oxytocin treatments on temporary social separation in adult titi monkeys (*Plecturocebus cupreus*). *Front Behav Neurosci*. 16: 877631.

Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE et al. (2019) Characterizing the major structural variant alleles of the human genome. *Cell*. 176(3): 663–675.

Babbi G, Martelli PL, Profiti G, Bovo S, Savojardo C, Casadio R (2017) eDGAR: a database of Disease-Gene Associations with annotated Relationships among genes. *BMC Genomics*. 18(Suppl 5): 554.

Bales KL, Mason WA, Catana C, Cherry SR, Mendoza SP (2007) Neural correlates of pair-bonding in a monogamous primate. *Brain Res*. 1184: 245–253.

Bales KL, Arias-del Razo R, Conklin QA, Hartman S, Mayer HS, Rogers FD et al. (2017) Titi monkeys as a novel non-human primate model for the neurobiology of pair bonding. *Yale J Biol Med*. 90(3): 373–387.

Bales KL, Ardekani CS, Baxter A, Karaskiewicz CL, Kuske JX, Lau AR et al. (2021) What is a pair bond? *Horm. Behav.* 136: 105062.

Basso C, Corrado D, Marcus FI, Nava A, Thiene G (2009) Arrhythmogenic right ventricular cardiomyopathy. *Lancet*. 373(9671): 1289–1300.

Belyeu JR, Brand H, Wang H, Zhao X, Pedersen BS, Feusier J et al. (2021) *De novo* structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *Am J Hum Genet*. 108(4): 597–607.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Borozdin W, Boehm D, Leipoldt M, Wilhelm C, Reardon W, Clayton-Smith J et al. (2004) SALL4 deletions are a common cause of Okihiro and acro-renal-ocular syndromes and confirm haploinsufficiency as the pathogenic mechanism. *J Med Genet*. 41(9): e113.

Brandler WM, Antaki D, Gujral M, Noor A, Rosanio G, Chapman TR et al. (2016) Frequency and complexity of *de novo* structural mutation in autism. *Am J Hum Genet*. 98(4): 667–679.

Brasó-Vives M, Povolotskaya IS, Hartasánchez DA, Farré X, Fernandez-Callejo M, Raveendran M et al. (2020) Copy number variants and fixed duplications among 198 rhesus macaques (*Macaca mulatta*). *PLoS Genet*. 16(5): e1008742.

Cameron DL, Schröder J, Penington JS, Do H, Molania R, Dobrovic A et al. (2017) GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res*. 27(12): 2050–2060.

Cerrone M, Montnach J, Lin X, Zhao YT, Zhang M, Agullo-Pascual E et al. (2017) Plakophilin-2 is required for transcription of genes that control calcium cycling and cardiac rhythm. *Nat Commun*. 8(1): 106.

Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T et al. (2019) Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun*. 10(1): 1784.

Carter CS, Kenkel WM, MacLean EL, Wilson SR, Perkeybile AM, Yee JR et al. (2020) Is oxytocin "nature's medicine"? *Pharmacol Rev*. 72(4): 829–861.

Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M et al. (2016) Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 32(8): 1220–1222.

Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP et al. (2015) SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat Methods*. 12(10): 966–968.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 6(2): 80–92.

Clarke CE, Alger S, Preece MA, Burdon MA, Chavda S, Denis S et al. (2004) Tremor and deep white matter changes in alpha-methylacyl-CoA racemase deficiency. *Neurology*. 63(1): 188–189.

Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC et al. (2020) A structural variation reference for medical and population genetics. *Nature*. 581(7809): 444–451.

Conrad DF, Hurles ME (2007) The population genetics of structural variation. *Nat Genet*. 39(7 Suppl): S30–S36.

Conrad DF, Pinto D, Redon R, Feuk L, Gökçümen O, Zhang Y et al. (2010) Origins and functional impact of copy number variation in the human genome. *Nature*. 464(7289): 704–712.

Corrado D, Link MS, Calkins H (2017) Arrhythmogenic right ventricular cardiomyopathy. *N Engl J Med*. 376(1): 61–72.

Dalal D, Molin LH, Piccini J, Tichnell C, James C, Bomma C et al. (2006) Clinical features of arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in plakophilin-2. *Circulation*. 113(13): 1641–1649.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience*. 10(2): giab008.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Delage WJ, Thevenon J, Lemaitre C (2020) Towards a better understanding of the low recall of insertion variants with short-read based variant callers. *BMC Genomics*. 21(1): 762.

Dennis MY, Harshman L, Nelson BJ, Penn O, Cantsilieris S, Huddleston J et al. (2017) The evolution and population diversity of human-specific segmental duplications. *Nat Ecol Evol*. 1(3): 69.

Dick D, Horvath R, Chinnery PF (2011) AMACR mutations cause late-onset autosomal recessive cerebellar ataxia. *Neurology*. 76(20): 1768–1770.

Drachman JG, Jarvik GP, Mehaffey MG (2000) Autosomal dominant thrombocytopenia: incomplete megakaryocyte differentiation and linkage to human chromosome 10. *Blood*. 96(1): 118–125.

Eggertsson HP, Kristmundsdottir S, Beyter D, Jonsson H, Skuladottir A, Hardarson MT et al. (2019) GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. *Nat Commun*. 10(1): 5402.

Emborg ME (2017) Nonhuman primate models of neurodegenerative disorders. *ILAR J*. 58(2): 190–201.

Ferdinandusse S, Denis S, Clayton PT, Graham A, Rees JE, Allen JT et al. (2000) Mutations in the gene encoding peroxisomal alpha-methylacyl-CoA racemase cause adult-onset sensory motor neuropathy. *Nat Genet*. 24(2): 188–191.

Feulner PGD, De-Kayne R (2017) Genome evolution, structural rearrangements and speciation. *J Evol Biol*. 30(8): 1488–1490.

Fischer D, Schabhüttl M, Wieland T, Windhager R, Strom TM, Auer-Grumbach M (2014). A novel missense mutation confirms ATL3 as a gene for hereditary sensory neuropathy type 1. *Brain*. 137(Pt 7): e286.

Gabrielaite M, Torp MH, Rasmussen MS, Andreu-Sánchez S, Vieira FG, Pedersen CB et al. (2021) A comparison of tools for copy-number variation detection in germline whole exome and whole genome sequencing data. *Cancers (Basel)*. 13(24): 6283.

Gerull B, Heuser A, Wichter T, Paul M, Basson CT, McDermott DA et al. (2004) Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat Genet*. 36(11): 1162–1164.

Gokcumen O, Tischler V, Tica J, Zhu Q, Iskow RC, Lee E et al. (2013) Primate genome architecture influences structural variation mechanisms and functional consequences. *Proc Natl Acad Sci U S A*. 110(39): 15764–15769.

Groves CP (2005) Species of *Callicebus* (*Callicebus*) *cupreus*. In: Wilson DE, Reeder DM, editors. Mammal species of the world: a taxonomic and geographic reference. 3rd ed. Baltimore: John Hopkins University Press. p. 142–143.

Harner MK, Bishop DV, Pollak RM, Purcell RH, Mulle JG (2025) Copy number variants: deletion and duplication syndromes. *Annu Rev Genomics Hum Genet*. 26(1): 261–277.

Hartasánchez DA, Brasó-Vives M, Heredia-Genestar JM, Pybus M, Navarro A (2018) Effect of collapsed duplications on diversity estimates: what to expect. *Genome Biol Evol*. 10(11): 2899–2905.

Hurley PJ, Elsworth JD, Whittaker MC, Roth RH, Redmond DE Jr (2011) Aged monkeys as a partial model for Parkinson's disease. *Pharmacol Biochem Behav*. 99(3): 324–332.

Ho SS, Urban AE, Mills RE (2020) Structural variation in the sequencing era. *Nat Rev Genet*. 21(3): 171–189.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Hollox EJ, Hoh BP (2014) Human gene copy number variation and infectious disease. *Hum Genet*. 133(10): 1217–1233.

Hollox EJ, Zuccherato LW, Tucci S (2022) Genome structural variation in human evolution. *Trends Genet*. 38(1): 45–58.

Horta M, Kaylor K, Feifel D, Ebner NC (2020) Chronic oxytocin administration as a tool for investigation and treatment: a cross-disciplinary systematic review. *Neurosci Biobehav Rev*. 108: 1–23.

Iskow RC, Gökçümen O, Lee C (2012) Exploring the role of copy number variants in human adaptation. *Trends Genet*. 28(6): 245–257.

Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ et al. (2010) *De novo* rates and selection of large copy number variation. *Genome Res*. 20(11): 1469–1481.

Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C et al. (2017) Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun*. 8: 14061.

Jennewein DM, Lee J, Kurtz C, Dizon W, Shaeffer I, Chapman A et al. (2023) The Sol Supercomputer at Arizona State University. In Practice and Experience in Advanced Research Computing 2023: Computing for the Common Good (PEARC '23). Association for Computing Machinery, New York, NY, USA, 296–301.

Kinzey W (1997) New World primates: ecology, evolution, and behavior. New York: Aldine de Gruyter.

Kirchner F, Schuetz A, Boldt LH, Martens K, Dittmar G, Haverkamp W et al. (2012) Molecular insights into arrhythmogenic right ventricular cardiomyopathy caused by plakophilin-2 missense mutations. *Circ Cardiovasc Genet*. 5(4): 400–411.

Kohlhase J, Heinrich M, Schubert L, Liebers M, Kispert A, Laccone F et al. (2002) Okihiro syndrome is caused by SALL4 mutations. *Hum Mol Genet*. 11(23): 2979–2987.

Kohlhase J, Schubert L, Liebers M, Rauch A, Becker K, Mohammed SN et al. (2003) Mutations at the SALL4 locus on chromosome 20 result in a range of clinically overlapping phenotypes, including Okihiro syndrome, Holt-Oram syndrome, acro-renal-ocular syndrome, and patients previously reported to represent thalidomide embryopathy. *J Med Genet*. 40(7): 473–478.

Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci*. 279(1749): 5048–5057.

Kornak U, Mademan I, Schinke M, Voigt M, Krawitz P, Hecht J et al. (2014) Sensory neuropathy with bone destruction due to a mutation in the membrane-shaping atlastin GTPase 3. *Brain*. 137(Pt 3): 683–692.

Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y (2019) Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol*. 20(1): 117.

Kosugi S, Terao C (2024) Comparative evaluation of SNVs, indels, and structural variations detected with short- and long-read sequencing data. *Hum Genome Var*. 11(1): 18.

Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, Meyerson OS et al. (2018) High-resolution comparative analysis of great ape genomes. *Science*. 360(6393): eaar6343.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S et al. (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 44(D1): D862–D868.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S et al. (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 46(D1): D1062–D1067.

Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 15(6): R84.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* [Preprint] https://doi.org/10.48550/arXiv.1303.3997

Li H, Su LY, Yang L, Li M, Liu Q, Li Z et al. (2020a) A cynomolgus monkey with naturally occurring Parkinson's disease. *Natl Sci Rev*. 8(3): nwaa292.

Li H, Wu S, Ma X, Li X, Cheng T, Chen Z et al. (2021) Co-editing PINK1 and DJ-1 genes via adeno-associated virus-delivered CRISPR/Cas9 system in adult monkey brain elicits classical Parkinsonian phenotype. *Neurosci Bull*. 37(9): 1271–1288.

Li J, Fan Z, Shen F, Pendleton AL, Song Y, Xing J et al. (2020b) Genomic copy number variation study of nine *Macaca* species provides new insights into their genetic divergence, adaptation, and biomedical application. *Genome Biol Evol*. 12(12): 2211–2230.

Liu Y, Zhang L, Jin X, Liu Z, Qi J, Fan Z et al. (2025) Genome-wide copy number variation drives adaptive evolution in *Macaca mulatta* populations. *Current Zoology*. zoaf050. Online ahead of print.

Lukas D, Clutton-Brock TH (2013) The evolution of social monogamy in mammals. *Science*. 341(6145): 526–530.

Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ (2019) Structural variant calling: the long and the short of it. *Genome Biol*. 20(1): 246.

Mao Y, Harvey WT, Porubsky D, Munson KM, Hoekzema K, Lewis AP et al. (2024) Structurally divergent and recurrently mutated regions of primate genomes. *Cell*. 187(6): 1547–1562.

Mendoza SP, Mason WA (1986) Parental division of labour and differentiation of attachments in a monogamous primate (*Callicebus cupreus*). *Anim Behav*. 34(5): 1336–1347.

Mertens C, Kuhn C, Franke WW (1996) Plakophilins 2a and 2b: constitutive proteins of dual location in the karyoplasm and the desmosomal plaque. *J Cell Biol*. 135(4): 1009–1025.

Miertus J, Borozdin W, Frecer V, Tonini G, Bertok S, Amoroso A et al. (2006) A SALL4 zinc finger missense mutation predicted to result in increased DNA binding affinity is associated with cranial midline defects and mild features of Okihiro syndrome. *Hum Genet*. 119(1-2): 154–161.

Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA et al. (2011) Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res*. 21(6): 830–839.

Moerkerke M, Daniels N, Tibermont L, Tang T, Evenepoel M, Van der Donck S et al. (2024) Chronic oxytocin administration stimulates the oxytocinergic system in children with autism. *Nat Commun*. 15(1): 58.

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A et al (2022) The complete sequence of a human genome. *Science*. 376(6588): 44–53.

O'Callaghan B, Hardy J, Plun-Favreau H (2023) PINK1: from Parkinson's disease to mitophagy and back again. *PLoS Biol*. 21(6): e3002196.

Ohta T (1993) An examination of the generation-time effect on molecular evolution. *Proc Natl Acad Sci U S A*. 90(22): 10676–10680.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Pan MT, Zhang H, Li XJ, Guo XY (2024) Genetically modified non-human primate models for research on neurodegenerative diseases. *Zool Res*. 45(2): 263–274.

Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF et al. (2010) Towards a comprehensive structural variation map of an individual human genome. *Genome Biol*. 11(5): R52.

Paradisi I, Arias S (2007) IVIC syndrome is caused by a c.2607delA mutation in the SALL4 locus. *Am J Med Genet A*. 143(4): 326–332.

Pfeifer SP (2017) From next-generation resequencing reads to a high-quality variant data set. *Heredity (Edinb)*. 118(2): 111–124.

Pfeifer SP, Baxter A, Savidge LE, Sedlazeck FJ, Bales KL (2024) *De novo* genome assembly for the coppery titi monkey (*Plecturocebus cupreus*): an emerging nonhuman primate model for behavioral research. *Genome Biol Evol*. 16(5): evae108.

Porubsky D, Eichler EE (2024) A 25-year odyssey of genomic technology advances and structural variant discovery. *Cell*. 187(5): 1024–1037.

Porubsky D, Sanders AD, Höps W, Hsieh P, Sulovari A, Li R et al. (2020) Recurrent inversion toggling and great ape genome evolution. *Nat Genet*. 52(8): 849–858.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 28(18): i333–i339.

Rigney N, de Vries GJ, Petrulis A, Young LJ (2022) Oxytocin, vasopressin, and social behavior: from neural circuits to clinical opportunities. *Endocrinology*. 163(9): bqac111.

Savoia A, Del Vecchio M, Totaro A, Perrotta S, Amendola G, Moretti A et al. (1999) An autosomal dominant thrombocytopenia gene maps to chromosomal region 10p. *Am J Hum Genet*. 65(5): 1401–1405.

Saxena AS, Baer CF (2025) High rate of mutation and efficient removal by selection of structural variants from natural populations of *Caenorhabditis elegans*. *BioRxiv* [Preprint]. doi: 10.1101/2025.03.22.644739.

Setchell KD, Heubi JE, Bove KE, O'Connell NC, Brewsaugh T, Steinberg SJ et al. (2003) Liver disease caused by failure to racemize trihydroxycholestanoic acid: gene mutation and effect of bile acid therapy. *Gastroenterology*. 124(1): 217–232.

Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC et al. (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*. 50(W1): W216–W221.

Smith EH, Gavrilov DK, Oglesbee D, Freeman WD, Vavra MW, Matern D et al. (2010) An adult onset case of alpha-methyl-acyl-CoA racemase deficiency. *J Inherit Metab Dis*. 33(Suppl 3): S349–S353.

Soni V, Pfeifer SP, Jensen JD (2025) Recent insights into the evolutionary genomics of the critically endangered aye-aye (*Daubentonia madagascariensis*). *Am J Primatol*. 87(12): e70105.

Subramanian K, Chopra M, Kahali B (2024) Landscape of genomic structural variations in Indian population-based cohorts: deeper insights into their prevalence and clinical relevance. *HGG Adv*. 5(3): 100285.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sudmant PH, Huddleston J, Catacchio CR, Malig M, Hillier LW, Baker C et al. (2013) Evolution and diversity of copy number variation in the great ape lineage. *Genome Res*. 23(9): 1373–1382.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J et al. (2015) An integrated map of structural variation in 2,504 human genomes. *Nature*. 526(7571): 75–81.

Taylor MS, Ponting CP, Copley RR (2004) Occurrence and consequences of coding sequence insertions and deletions in mammalian genomes. *Genome Res*. 14(4): 555–566.

Terbot JW, Soni V, Versoza CJ, Bales KL, Pfeifer SP, Jensen JD. 2026. Inferring the demographic history of coppery titi monkeys (*Plecturocebus cupreus*) from high-quality, whole-genome, population-level data. *BioRxiv* [preprint]. doi: 10.64898/2026.01.09.698678.

Thiene G, Nava A, Corrado D, Rossi L, Pennelli N (1988) Right ventricular cardiomyopathy and sudden death in young people. *N Engl J Med*. 318(3): 129–133.

Thomas GWC, Wang RJ, Nguyen J, Harris RA, Raveendran M, Rogers J et al. (2021) Origins and long-term patterns of copy-number variation in rhesus macaques. *Mol Biol Evol*. 38(4): 1460–1471.

Thompson SA, Calvin J, Hogg S, Ferdinandusse S, Wanders RJ, Barker RA (2008) Relapsing encephalopathy in a patient with alpha-methylacyl-CoA racemase deficiency. *J Neurol Neurosurg Psychiatry*. 79(4): 448–450.

UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res*. 43(Database issue): D204–D212.

van der Auwera GA, O'Connor BD (2020) Genomics in the cloud: using Docker, GATK, and WDL in Terra. Sebastopol: O'Reilly Media.

Valeggia CR, Mendoza SP, Fernandez-Duque E, Mason WA, Lasley B (1999) Reproductive biology of female titi monkeys (*Callicebus moloch*) in captivity. *Am J Primatol*. 47(3): 183–195.

van Tintelen JP, Entius MM, Bhuiyan ZA, Jongbloed R, Wiesfeld AC, Wilde AA et al. (2006) Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circulation*. 113(13): 1650–1658.

Versoza CJ, Jensen JD, Pfeifer SP (2025) The landscape of structural variation in aye-ayes (*Daubentonia madagascariensis*). *Genome Biol Evol*. 17(9): evaf167.

Witczak LR, Samra J, Dufek M, Goetze LR, Freeman SM, Lau AR et al. (2024) Expression of bond-related behaviors affects titi monkey responsiveness to oxytocin and vasopressin treatments. *Ann N Y Acad Sci*. 1534(1): 118–129.

Wittinghofer A, Vetter IR (2011) Structure-function relationships of the G domain, a canonical switch motif. *Annu Rev Biochem*. 80:943–971.

Wold J, Koepfli KP, Galla SJ, Eccles D, Hogg CJ, Le Lec MF et al. (2021) Expanding the conservation genomics toolbox: incorporating structural variants to enhance genomic studies for species of conservation concern. *Mol Ecol*. 30(23): 5949–5965.

Wold JR, Guhlin JG, Dearden PK, Santure AW, Steeves TE (2025) The promise and challenges of characterizing genome-wide structural variants: a case study in a critically endangered parrot. *Mol Ecol Resour*. 25(5): e13783.

Xu H, Zhang C, Cao L, Song J, Xu X, Zhang B et al. (2019) ATL3 gene mutation in a Chinese family with hereditary sensory neuropathy type 1F. *J Peripher Nerv Syst*. 24(1): 150–155.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Yang W, Liu Y, Tu Z, Xiao C, Yan S, Ma X et al. (2019) CRISPR/Cas9-mediated PINK1 deletion leads to neurodegeneration in rhesus monkeys. *Cell Res*. 29(4): 334–336.

Yang W, Guo X, Tu Z, Chen X, Han R, Liu Y et al. (2022) PINK1 kinase dysfunction triggers neurodegeneration in the primate brain without impacting mitochondrial homeostasis. *Protein Cell*. 13(1): 26–46.

Yang X, Mao Y, Wang XK, Ma DN, Xu Z, Gong N et al. (2023) Population genetics of marmosets in Asian primate research centers and loci associated with epileptic risk revealed by whole-genome sequencing. *Zool Res*. 44(5): 837–847.

Yang Y, Braga MV, Dean MD (2024) Insertion-deletion events are depleted in protein regions with predicted secondary structure. *Genome Biol Evol*. 16(5): evae093.

Yin EP, Dieriks BV (2025) Rethinking 'rare' *PINK1* Parkinson's disease: a meta-analysis of geographical prevalence, phenotypic diversity, and α-synuclein pathology. *J Parkinsons Dis*. 15(1): 41–65.

Young E, Abid HZ, Kwok PY, Riethman H, Xiao M (2020) Comprehensive analysis of human subtelomeres by whole genome mapping. *PLoS Genet*. 16(1): e1008347.

Zablocki-Thomas P, Lau A, Witczak L, Dufek M, Wright A, Savidge L et al. (2023) Intranasal oxytocin does not change partner preference in female titi monkeys (*Plecturocebus cupreus*), but intranasal vasopressin decreases it. *J Neuroendocrinol*. 35(10): e13339.

Zhang S, Xu N, Lu Y, Nie Y, Li Z, de Gennaro L et al. (2025) A complete and near-perfect rhesus macaque reference genome: lessons from subtelomeric repeats and sequencing bias. *BioRxiv* [Preprint]. doi: 10.1101/2025.08.04.668424.

Table 1. Structural variants with major effects predicted to affect disease-linked genes.

type	chr	start	size	predicted effect	allele freq.	human gene	human disease phenotype
DEL	1	10,718,272	1,041	exon loss variant splice acceptor/donor variant	0.25	ATL3	hereditary sensory & autonomic neuropathy
	1	136,329,129	1,569	frameshift variant splice acceptor variant	0.42	ANKRD26	thrombocytopenia
	1	167,808,572	68	frameshift variant start lost	0.15	PKP2	arrhythmogenic cardiomyopathy
	1	183,000,971	534	frameshift variant splice donor variant	0.06	SALL4	IVIC syndrome Duane-radial ray syndrome
INV	1	201,462,663	7,384	splice acceptor variant	0.12	AMACR	AMACR deficiency congenital bile acid synthesis defect
	13	99,538,214	1,950	splice acceptor/donor variant	0.04	PINK1	Parkinson's disease (early-onset)

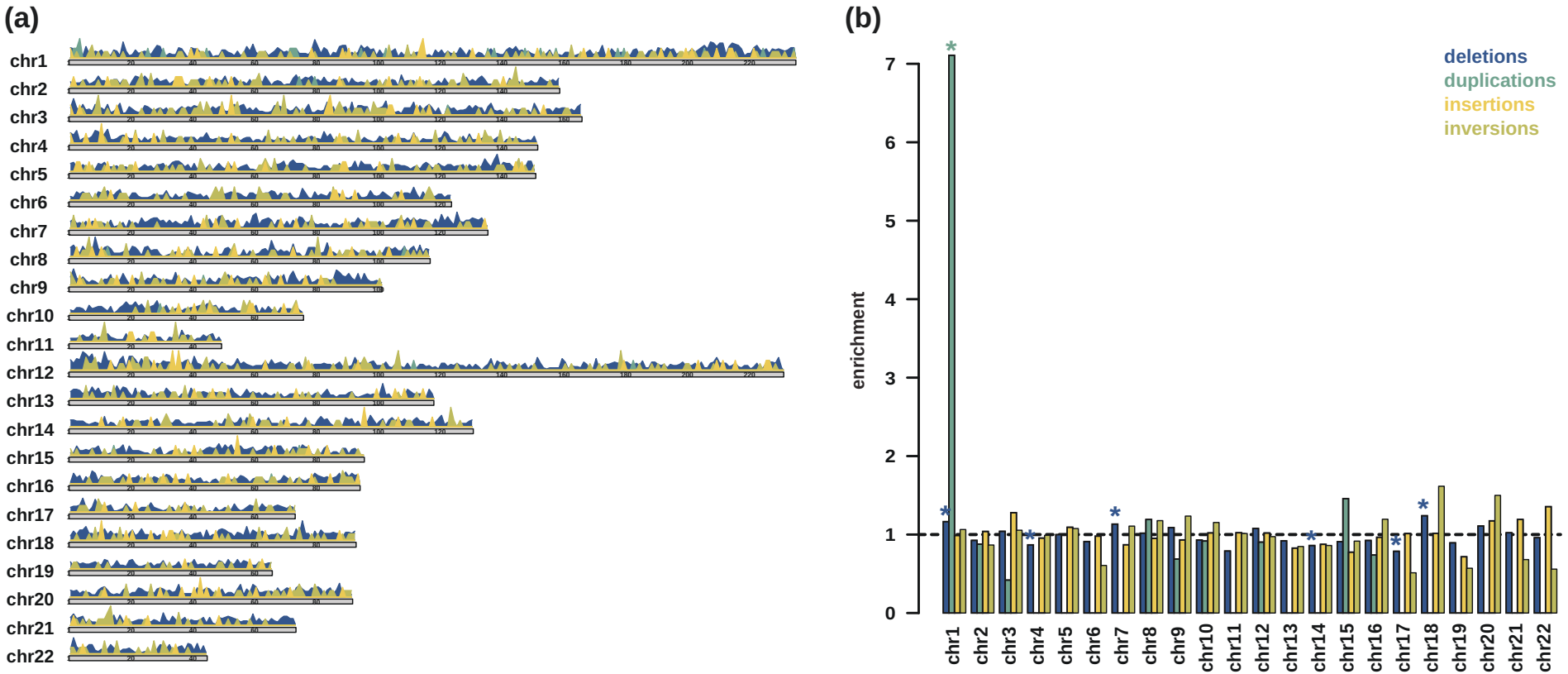

Figure 1. **Landscape of structural variation in the coppery titi monkey genome**. (a) Map of autosomal copy number variants (with deletions color-coded in blue and duplications in teal), insertions (yellow) and inversions (olive green). Chromosomes are displayed as horizontal bars and variants as individual peaks, with the height of each peak being proportional to the region length. (b) Distribution of structural variants across the autosomes. Significant enrichments/depletions are marked by a star (*).

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

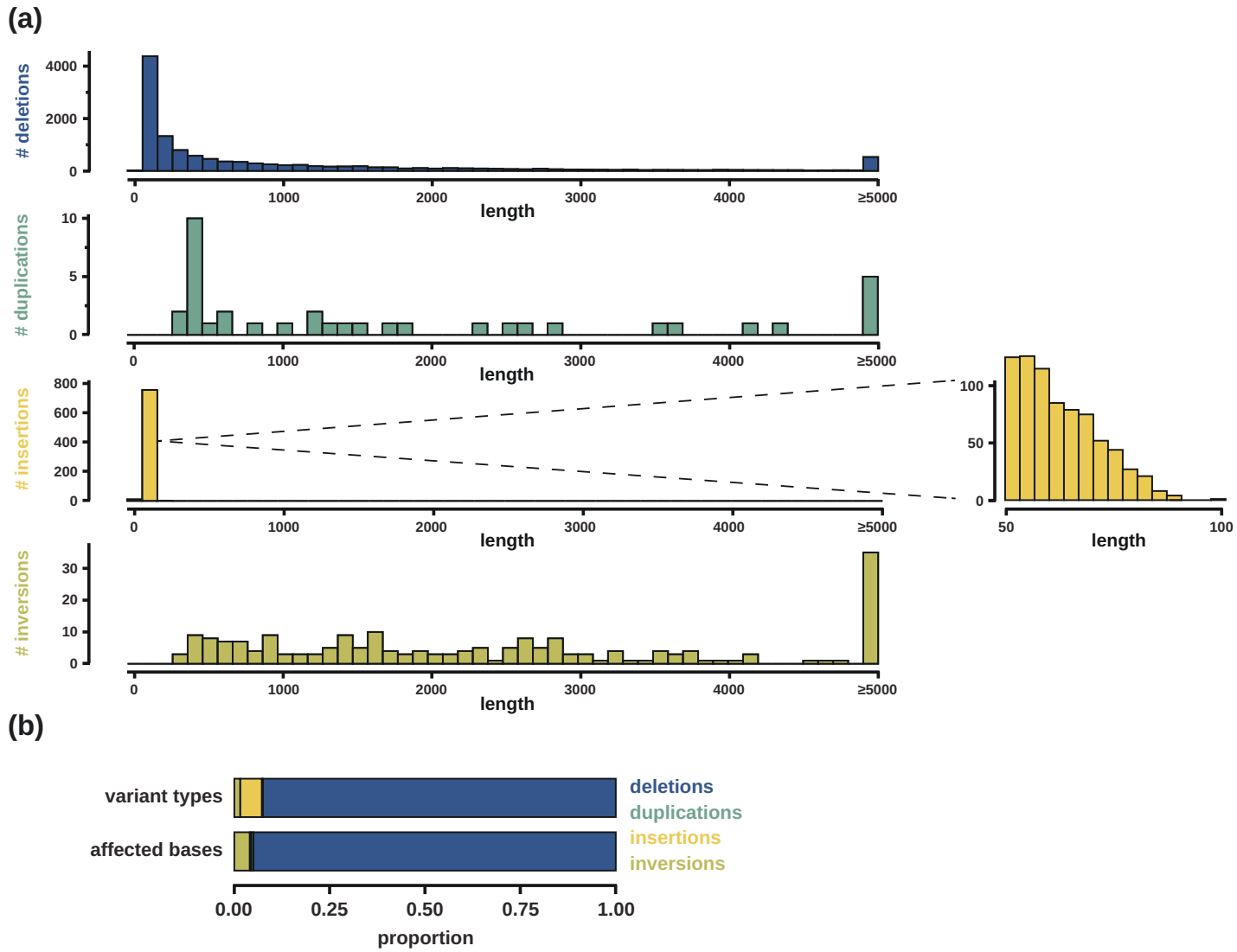

Figure 2. **Characteristics of structural variation in the coppery titi monkey genome**. (a) Length distribution of autosomal copy number variation (with deletions color-coded in blue and duplications in teal), insertions (yellow) and inversions (olive green). (b) Proportion of the base-pairs affected by the different variant types.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.13.699302; this version posted January 14, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

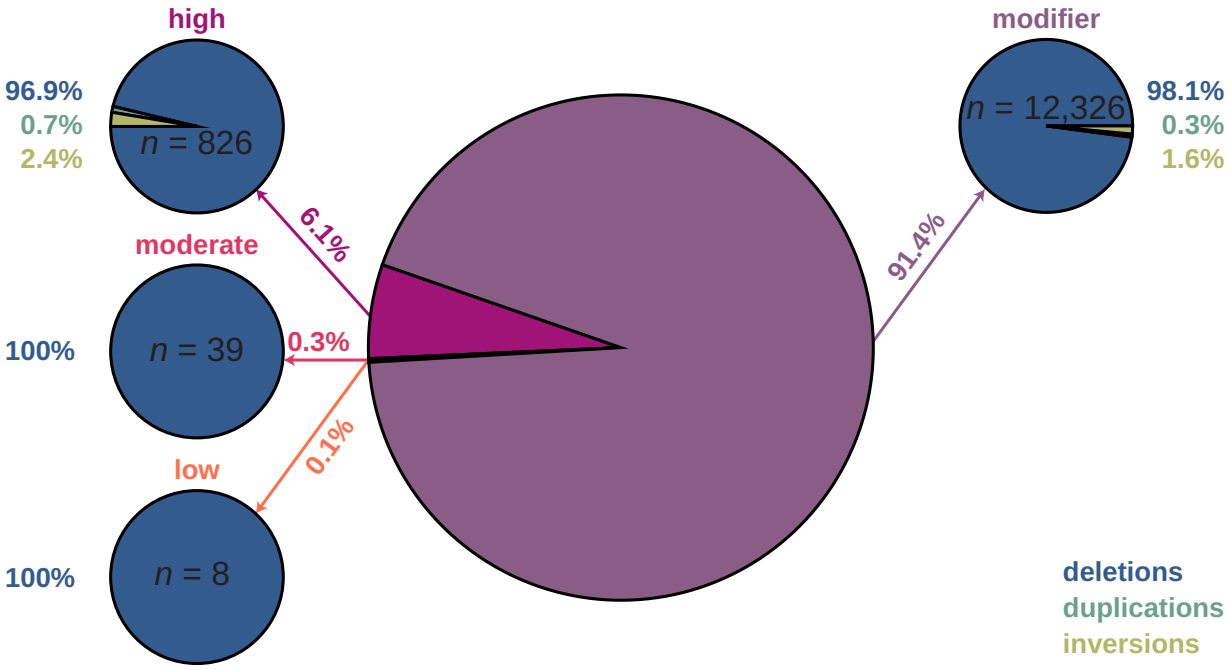

Figure 3. **Functional annotation of structural variation in the coppery titi monkey genome**. The proportion of autosomal copy number variation (with deletions color-coded in blue and duplications in teal), insertions (yellow) and inversions (olive green) predicted within non-coding or regulatory regions (i.e., modifiers; shown in purple) as well as those predicted to be likely disruptive (high: pink), more subtly affect protein function or gene regulation (moderate: rose), and unlikely to have a major functional impact (low: orange). Note that due to their size, several structural variants were predicted to have multiple effects.