Evolution and codon usage bias of mitochondrial and nuclear genomes in *Aspergillus* section *Flavi*

Miya Hugaboom[1], E. Anne Hatmaker[1,2,*], Abigail L. LaBella[3], and Antonis Rokas[1,2,*]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN, 37235, USA

[2] Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN, 37235, USA

[3]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, 28223, USA

*Authors for correspondence: e.anne.hatmaker@vanderbilt.edu & antonis.rokas@vanderbilt.edu

Author ORCiDs:

Miya Hugaboom: https://orcid.org/0000-0001-8099-3564

E. Anne Hatmaker: https://orcid.org/0000-0002-7821-8160

Abigail L. LaBella: https://orcid.org/0000-0003-0068-6703

Antonis Rokas: https://orcid.org/0000-0002-7248-6551

Running head: mtDNA evolution in *Aspergillus*

Mailing address: VU Station B 35-1364, Nashville, TN 37235, U.S.A.

Keywords: mitogenome, phylogeny, codon usage bias, *Aspergillus*, section *Flavi*, fungi

Abstract

The fungal genus *Aspergillus* contains a diversity of species divided into taxonomic sections of closely related species. Section *Flavi* contains 33 species, many of industrial, agricultural, or medical relevance. Here, we analyze the mitochondrial genomes (mitogenomes) of 20 *Flavi* species—including 18 newly assembled mitogenomes—and compare their evolutionary history and codon usage bias (CUB) patterns to their nuclear counterparts. CUB refers to variable frequencies of synonymous codons in coding DNA and is shaped by a balance of neutral processes and natural selection. All mitogenomes were circular DNA molecules with highly conserved gene content and order. As expected, genomic content, including GC content, and genome size differed greatly between mitochondrial and nuclear genomes. Phylogenetic analysis based on 14 concatenated mitochondrial genes predicted evolutionary relationships largely consistent with those predicted by a phylogeny constructed from 2,422 nuclear genes. Comparing similarities in interspecies patterns of CUB between mitochondrial and nuclear genomes showed that species grouped differently by patterns of CUB depending on whether analyses were performed using mitochondrial or nuclear relative synonymous usage values. We found that patterns of CUB at gene-level are more similar between mitogenomes of different species than the mitogenome and nuclear genome of the same species. Finally, we inferred that, although most genes—both nuclear and mitochondrial—deviated from the neutral expectation for codon usage, mitogenomes were not under translational selection while nuclear genomes were under moderate translational selection. These results contribute to the study of mitochondrial genome evolution in filamentous fungi.

Introduction

The fungal genus *Aspergillus* is an important genus of filamentous fungi. The genus houses species with industrial applications, important pathogens of humans, animals and crops, producers of potent carcinogenic mycotoxins, and the genetic model organism *Aspergillus nidulans* (de Vries et al. 2017). *Aspergillus* is divided into taxonomic sections of closely related species. Section *Flavi* consists of 33 species, many of which have industrial, agricultural, or medical relevance (Gourama and Bullerman 1995; Hedayati et al. 2007; de Vries et al. 2017; Frisvad et al. 2018; Homa et al. 2019). For example, *A. oryzae* constitutes an important cell factory for enzyme production and, along with *A. sojae*, is vital to the production of a range of fermented foods (Machida et al. 2008; Sato et al. 2011). Conversely, *A. flavus* is an effective producer of aflatoxin B, a potent carcinogenic mycotoxin, and has been found to be both a plant contaminant and occasional pathogen, as well as an opportunistic human pathogen (Hedayati et al. 2007; Hoffmeister and Keller 2007; Dolezal et al. 2014). To better understand the diversity of these fungi, a recent study sequenced the genomes for 23 of the 33 known *Flavi* species to gain insights into their biology and evolution (Kjærbølling et al. 2020).

Previous genomic analyses of section *Flavi* focus almost exclusively on the nuclear genomes of the sequenced species (de Vries et al. 2017; Kjærbølling et al. 2020); the sole exception was a 2012 study that described the genomes of six diverse *Aspergillus* species, including two from section *Flavi* (Joardar et al. 2012). However, whole genome sequencing captures nucleotide sequences from both nuclear and organellar genomes. Fungal mitochondria have been linked to diverse processes including energy metabolism, cell differentiation, drug resistance, biofilm and hyphal growth regulation, and virulence, amongst others (Sanglard et al. 2001; Burger et al.

2003; Martins et al. 2011; Chatre and Ricchetti 2014; Calderone et al. 2015). Using appropriate software, mitochondrial reads can be effectively filtered and separated from nuclear reads within existing whole-genome sequencing datasets to be used for mitochondrial genome (mitogenome) assembly and annotation (Hugaboom et al. 2021). Fungal mitogenomes, including those of *Aspergillus* species, are typically circular and composed of a single chromosome (Brown et al. 1985; Joardar et al. 2012). Mitogenomes replicate independently from the nuclear genome and cell cycle and tend to have high copy number. Fourteen protein-coding genes involved in the electron transport chain are highly conserved within fungal mitogenomes (Gray et al. 1999; Lavín et al. 2008; Joardar et al. 2012). Genes for two ribosomal rRNAs subunits, one large and one small, and a variable number of tRNAs also tend to be housed in the mitogenome (Gray et al. 1999; Lavín et al. 2008; Joardar et al. 2012). Variation in fungal mitogenomes is largely due to differences in intron distribution and the variable presence of accessory mitochondrial genes, even between closely related species (Joardar et al. 2012; Li, Xiang, et al. 2019; Li, Wang, et al. 2019; Wang et al. 2020; Zhang et al. 2020; Chen et al. 2021). Importantly, mitogenomes also differ from nuclear genomes in their inheritance pattern. Although fungal mitogenomes are not always uniparentally inherited and can exhibit recombination (Basse 2010; Stein and Sia 2017; Zardoya 2020; Mukhopadhyay and Hausner 2021), *Aspergillus* mitogenomes are uniparentally inherited and rarely display recombination, offering a unique phylogenetic perspective (Coenen et al. 1996; Santamaria et al. 2009; Kjærbølling et al. 2020). Mitochondrial genomes may therefore hold clues to both the biology and evolution of these fungal species.

Another key difference between mitogenomes and nuclear genomes is codon usage bias (CUB). CUB refers to the different frequency of synonymous codons—those that code for the same

amino acid—in coding DNA. Changes in synonymous codons do not alter primary protein sequence and were thus once assumed to be selectively neutral (Jia and Higgs 2008; Wei et al. 2014; LaBella et al. 2019). However, CUB has been found to influence numerous cellular processes, particularly those associated with translation (Stoletzki and Eyre-Walker 2007; Zhou et al. 2009). This is hypothesized to be due to codon optimization: the tendency for codon usage to be correlated to the abundance of tRNA molecules in the genome (Post et al. 1979; Nakamura et al. 1980; Ikemura 1981; Gouy and Gautier 1982; P M Sharp and Li 1986; Thomas et al. 1988). During translation, mRNAs containing optimized codons—codons corresponding to the tRNA pool of the cell—are translated more efficiently than those with non-optimal codon usage (Bulmer 1991; Xia 1998; Chevance et al. 2014; Presnyak et al. 2015; Hanson and Coller 2018). In many organisms, this leads to a correlation between codon usage and protein production (Ikemura 1981; Bulmer 1991; Gustafsson et al. 2004; Hiraoka et al. 2009; Roymondal et al. 2009; Zhipeng et al. 2016; Payne and Alvarez-Ponce 2019; Sahoo et al. 2019). Importantly, mitogenomes house their own set of tRNAs that is distinct from that of the nuclear genome and thus may exhibit patterns of CUB shaped by optimization to a greater extent by the mitochondrial set of tRNAs (tRNAome) than the nuclear tRNAome. Variation in synonymous codon usage is a widespread phenomenon at codon, gene, and whole genome levels in nuclear and mitochondrial genomes (LaBella et al. 2019; LaBella et al. 2021; Wint et al. 2022). This variation in codon usage likely reflects a balance of mutational bias (e.g., GC content), natural selection (e.g., translational selection), and genetic drift (Ikemura 1985; Shields and Sharp 1987; Sharp et al. 1993; Wei et al. 2014). The balance of these forces varies between organisms. In many microbes, for example, translational selection plays a large role, whereas mutational bias plays the primary role in humans (Sharp et al. 1993). However, analysis of mitochondrial CUB

in fungi is limited (Kamatani and Yamamoto 2007; Carullo and Xia 2008). Understanding patterns of CUB can provide insight into the evolutionary history of individual genes and entire genomes.

To gain insights into the evolution of mitogenomes from the section *Flavi,* we analyzed mitochondrial genomes of 20 section *Flavi* species—including 18 newly assembled ones—and compared their phylogeny and CUB to the nuclear genomes of the same species. All mitogenomes were confirmed to be circular DNA molecules of low GC content with highly conserved gene content and gene order. Genomic content and size differed greatly between mitochondrial and nuclear genomes. We then inferred and compared phylogenies constructed from mitochondrial versus nuclear genes. The presence and high copy number of mitogenomes within the cell as well as the lack of recombination relative to nuclear genomes (for a discussion of fungal mitochondrial genome recombination, see Zardoya 2020, Mukhopadhyay and Hausner 2021, and Stein and Sia 2017) make mitochondrial genes and genomes useful markers for phylogenetic analyses. Currently, phylogenies constructed for *Aspergillus* section *Flavi* are based solely on nuclear genome markers (Kjærbølling et al. 2020; Shen et al. 2020). Phylogenetic analysis based on 14 concatenated mitochondrial genes (mitogenes) predicted evolutionary relationships largely consistent with those inferred by a phylogeny based on nuclear data. We then investigated CUB in mitochondrial and nuclear genomes. At the gene-level, we found that patterns of CUB reflect whether the gene is mitochondrial or nuclear in origin as well as mitogene identity rather than species of origin; these patterns were influenced largely by GC content of the third codon position. Finally, we determined that although most genes—both nuclear and mitochondrial—deviated from the neutral expectation, mitogenomes were not under

translational selection while nuclear genomes were under moderate translational selection. By providing mitogenome assemblies for 20 section *Flavi* species and comparing the evolution of mitochondrial and nuclear genes in section *Flavi*, our study advances our understanding of genome evolution in the genus *Aspergillus*.

Methods

Genomic Data

We used strains from 21 species within section *Flavi* and *Aspergillus niger* (section *Nigri*) as an outgroup for phylogenetic analyses. For the mitochondrial dataset, we used a combination of available mitochondrial reference genomes and newly assembled whole-genome sequencing reads. Three previously assembled mitochondrial reference genomes (*Aspergillus sojae*, *Aspergillus oryzae,* and *Aspergillus niger*) were downloaded from NCBI's Nucleotide Database (Juhász et al. 2008; Machida et al. 2008; Sato et al. 2011). For new assemblies, previously sequenced paired-end Illumina whole genome sequence reads were downloaded from NCBI's Sequence Read Archive (Kjærbølling et al. 2020; Hatmaker et al. 2022).

Annotated protein-coding nucleotide sequences (CDS) for each nuclear genome were downloaded from JGI MycoCosm (Grigoriev et al. 2014; Kjærbølling et al. 2020) except for *A. sojae, A. flavus, and A. nomiae.* For *A. sojae,* strain-matched nuclear annotations were not available and thus this species was not included in phylogenetic inferences or any analyses based on nuclear genomic data. For *A. flavus* and *A. nomiae*, we used annotations from recently assembled genomes (Hatmaker et al., 2022), extracting the CDS regions. Strains and data sources are summarized in Table 1.

Mitochondrial Genome Assembly

Data from the whole genome sequence read files were extracted into usable format (FASTQ files) using SRA Toolkit v2.9.6-1 (Leinonen et al. 2011). Mitochondrial genomes were assembled from the raw reads of each species using the organelle genome assembler GetOrganelle v1.7.4.1 (Jin et al. 2020). Following the method of Hugaboom et al. (2021), we used the internal GetOrganelle fungal database (-F fungus_mt) and default parameter values for number of threads, extension, and k-mers to assemble the mitogenomes (Hugaboom et al. 2021). The complete mitochondrial genome for *Aspergillus fumigatus* SGAir0713 (GenBank accession: CM16889.1) was used as a reference for the seed database (parameter -s) for mitogenome assembly. Contigs generated for each *Aspergillus* species were circularized such that there was no overlap in the beginning and end of the mitochondrial genome sequence.

Read Mapping

Read mapping to correct errors was carried out using Bowtie2 v2.3.4.1 (Langmead and Salzberg 2012) and SAMtools v1.6 (Li et al. 2009). Bowtie2 aligned the raw paired-end reads from each *Aspergillus* species against the corresponding circularized mitochondrial genome. Variants were identified using SAMtools. Read mapping was also visualized and variants identified using the Integrative Genomics Viewer (IGV) v2.9.4 (Robinson et al. 2017).

Mitogenome Annotation

The rapid organellar genome annotation software GeSeq v2.03 (Tillich et al. 2017) was used to annotate the circularized mitochondrial genomes. In addition to the newly assembled

mitogenomes, mitochondrial reference genomes for *A. oryzae* and *A. sojae* were also annotated using GeSeq. Gene names were adjusted, and translations were checked in accordance with the reference mitochondrial genomes of *A. flavus* TCM2014 (NC_026920.1), *A. oryzae* 3.042 (NC_018100.1), *A. parasiticus* (NC_041445.1), and *A. fumigatus* A1163 (NC_017016.1). For *rnl* genes, GeSeq output was adjusted in accordance with both manual inspection in comparison to the above reference mitogenomes and NCBI Blast for similar sequences. Annotations were finalized following inspection of automated gene sequences using Geneious Prime v2021.1 (Kearse et al. 2012). OGDraw v1.1.1 (Greiner et al. 2019) was used for genome visualization.

Multiple Sequence Alignment

Using MAFFT v7 (Katoh et al. 2019), single-gene multiple sequence alignment (MSA) files based on DNA nucleotide sequences were created for each of the 14 core mitogenes: cytochrome oxidase subunits 1, 2, and 3, NADH dehydrogenase subunits 1, 2, 3, 4, 4L, 5, and 6, ATP synthase subunits 6, 8, and 9, and cytochrome b. Gene nucleotide sequences corresponding to translated amino acid sequences for each gene were extracted from Geneious Prime v2021.1 (Kearse et al. 2012) sequence view and reverse complemented as necessary. The 14 individual MSA files were concatenated using SequenceMatrix v1.9 (Vaidya et al. 2011).

Phylogenetic inference

To infer the evolutionary relationships within section *Flavi,* maximum likelihood phylogenies were constructed from both mitochondrial and nuclear data. The mitochondrial phylogeny was constructed from the MSA of 14 core concatenated mitogene nucleotide sequences files using RAxML v8.2.11 (Stamatakis 2014). The MSA was trimmed with ClipKIT v1.3.0 (Steenwyk et

al. 2020) to retain parsimony-informative sites prior to construction of the phylogeny. *A. niger* (NC_007445.1) was used as the outgroup. We used 1,000 bootstrap replicates to evaluate robustness of inference. For the nuclear phylogeny, orthologous proteins in all species were identified using OrthoFinder v2.5.4 (Emms and Kelly 2019). MSAs for each of 2,422 orthologs were concatenated using the script catfasta2phyml.pl (https://github.com/nylander/catfasta2phyml). The maximum likelihood nuclear phylogeny was constructed with 1,000 replicates for bootstrapping using RAxML v8.2.11 (Stamatakis 2014) from the aligned orthologs shared among all the *Aspergillus* species in the study (including *A. niger*) except *A. sojae*, which did not have available sequencing data for nuclear genome assembly and annotation. For both nuclear and mitochondrial phylogenies, GTR + Γ substitution models were used in accordance with model testing performed within the raxmlGUI 2.0 platform (Kozlov et al. 2019; Edler et al. 2021). The resulting consensus trees for both the mitochondrial and nuclear phylogenies were visualized using Geneious Prime v2020.1.2 (Kearse et al. 2012).

Cluster Analysis

To compare patterns of synonymous codon usage bias between mitochondrial and nuclear genomes, hierarchical clustering of genome-level relative synonymous codon usage (RSCU) values was calculated and visualized using RStudio v. 2021.09.1. RSCU is a commonly used metric for codon usage bias that reflects the observed frequency of a particular codon divided by its expected frequency if all synonymous codons were used equally (Paul M Sharp and Li 1986). Genome-level RSCU values as well as RSCU values for each mitochondrial and nuclear gene were computed using DAMBE v7.3.5 (Xia 2017).

Correspondence Analysis

To determine which codons drive differences in signatures of codon usage between nuclear and mitochondrial genes and between the mitogenomes of the 20 *Flavi* species, correspondence analyses were performed using gene-level RSCU values. Correspondence analysis was used for multivariate analysis because the RSCU values are interdependent—the RSCU values for one codon are inherently linked to the RSCU values of other synonymous codons—and thus not suited for principal component analysis. The correspondence analyses were carried out in RStudio v. 2021.09.1. using the packages ade4 v.1.7-19(https://CRAN.R-project.org/package=ade4) and factoextra v.1.0.7 (https://CRAN.R-project.org/package=factoextra).

Evaluation of mutational bias and codon usage

To evaluate the role of mutational bias in determining the observed patterns of codon usage bias in section *Flavi,* we plotted the effective number of codons (ENc) for each gene against their respective GC3 values, where GC3 is the GC content of the third codon position. ENc is often used to assess the non-uniformity of synonymous codon usage within individual genes (Wright 1990). Values range from 20 (extreme bias where only one codon is used per amino acid) to 61 (no bias). The ENc values for each gene were computed in DAMBE v.7.3.5 (Xia 2017). The resulting distribution was compared to the predicted neutral distribution proposed by dos Reis et al. (dos Reis et al. 2004) using the suggested parameters by computing the R^2 values between the observed and expected ENc values.

Evaluation of selection on codon usage

253 To compare the influence of translational selection on the codon usage bias of mitogenomes
254 as compared to nuclear genomes, we calculated the S-value proposed by dos Reis et al. (2004)
255 for each species. The S-value is the correlation between the tRNA adaptation index (stAI) and
256 the confounded effects of selection on the codon usage of a gene as well as of other factors (e.g.,
257 mutation bias, genetic drift). Therefore, the S-value measures the proportion of the variance in
258 codon bias that cannot be accounted for without invoking translational selection. Thus, the
259 higher the S-value, the stronger the action of translational selection on the given set of genes.
260
261 To calculate the S-value, we first measured tRNA counts for each nuclear and mitochondrial
262 genome using tRNAscan-SE 2.0 (Chan et al. 2021). These counts were used to calculate the
263 species-specific value for each codon's relative adaptiveness (wi) in stAIcalc, version 1.0 (Sabi
264 et al. 2017). Exclusively mitochondrial tRNA counts were used to obtain wi values for
265 mitogenomes, whereas exclusively nuclear genome tRNA counts were used for nuclear
266 genomes. Taking the geometric mean of all wi values for the codons yielded the stAI of each
267 gene. These stAI values were then used to calculate S-values for each mitochondrial and nuclear
268 genome with the R package tAI.R, version 0.2 (https://github.com/mariodosreis/tai).
269
270 The statistical significance of each S-value was tested via a permutation test. 100 permutations
271 were run such that each genome's wi values were randomly assigned to codons, the tAI values
272 recalculated for each gene, and the S-test run on that permutation. A genome's observed S-value
273 was considered statistically significant if it fell in the top 5% of the distribution formed by the
274 100 values obtained by the permutation analysis.
275

Results

Genomic content varies greatly between nuclear and mitochondrial genomes

All mitogenomes were found to be small, circular DNA molecules with low GC content of 24.9-26.9% (Table 2, Figure 1). Each mitogenome contained fourteen core genes (cytochrome oxidase subunits 1, 2, and 3, NADH dehydrogenase subunits 1, 2, 3, 4, 4L, 5, and 6, ATP synthase subunits 6, 8, and 9, and cytochrome b) with conserved order and shared synteny (Figure 2). Additionally, a ribosomal protein S3 was found in all newly annotated *Flavi* genomes, and an intron encoded LAGLIDADG endonuclease was found in all mitogenomes except for *A. avenaceus* and *A. leporis.* Variations in mitogenome length are due to variations in intron number and length, primarily in the *cox1* gene. Introns were universally present in the *cox1* gene, with most mitogenomes housing a single intron ranging from 1,393-1,780 bp. Two exceptions—*A. avenaceus* and *A. coremiiformis*—housed 3 and 4 introns of total length 4,344 and 3,504 bp, respectively, in their *cox1* genes. All mitogenomes also housed a single intron in their *rnl* gene ranging from 1,682-1,709 bp. Finally, *A. avenaceus* and *A. coremiiformis* were found to have additional introns. *A. avenaceus* has a 1,227 bp intron in its *atp9* gene and a 1,200 bp intron in its *cob* gene, while *A. coremiiformis* has a 1,104 bp intron in its *nad5* gene and a 1,379 bp intron in its *cob* gene.

Conversely, corresponding nuclear genomes are linear and have less extreme GC content biases ranging from 43.0-48.8% (Table 2). Nuclear genomes are roughly 1,000 times larger than their mitochondrial counterparts; while mitogenomes ranged from 29.100-39.269 Kbp, nuclear genomes ranged from 30,1001-40,900 Kbp. Of note, both nuclear and mitochondrial genomes house their own set of tRNAs (i.e., have their own tRNAome), although the tRNAome of nuclear genomes is roughly ten times larger than that of mitochondrial genomes. While nuclear genomes

house 228-272 tRNAs, the mitogenomes encode a conserved set of 26 tRNAs. Importantly, each amino acid is represented by at least one tRNA in the conserved mitochondrial tRNAome. However, the codons GCC, GCU, CGG, CUC, CUU, CCC, CCU, UCC, UCG, UCU, ACC, ACU, GUC, GUU, and UGG could not be decoded without invoking additional wobble hypotheses, modification of mitochondrial tRNAs or importation of nuclear tRNAs.

Mitochondrial and nuclear phylogenies are very similar

To understand how the evolutionary history of section *Flavi* informed by mitogenomes compares to that of nuclear genomes, a mitochondrial phylogeny was constructed using a concatenation of 14 core mitogene nucleotide sequences (Figure 3B). The resulting phylogeny displayed high bootstrap support. Despite minor topological differences from a well-supported nuclear phylogeny (Figure 3A) amongst more closely related species, the evolutionary relationships predicted by the mitochondrial phylogeny largely align with those predicted by the nuclear phylogeny. For instance, although *A. minisclerotigenes, A. sergii, A. flavus, A. arachidicola, A. parasiticus, and A. novoparasiticus* fall within the same clade in both nuclear and mitochondrial phylogenies, the predicted evolutionary relationships within this clade vary slightly. Evolutionary rate was found to be more rapid in mitochondrial genomes relative to nuclear genomes. For example, the evolutionary distance between *A. flavus* and *A. nomiae* was 0.086 substitutions per site in the nuclear phylogeny, but 0.244 substitutions per site in the mitochondrial phylogeny. Single-gene mitochondrial phylogenies differed in their topologies but exhibited low bootstrap support values, particularly for relationships among closely related species (Supplementary File S1).

Species groupings based on patterns of codon usage bias differ between mitochondrial and nuclear genomes

To compare similarities in interspecies patterns of CUB between mitochondrial and nuclear genomes, hierarchical clustering was performed using the net RSCU values of protein-coding regions of both nuclear (Figure 4A) and mitochondrial (Figure 4B) genomes. The cluster analyses predict different interspecies relationships depending on organelle of genomic origin. For example, the cluster dendrograms show that *A. coremiiformis* and *A. avenaceus* cluster together based on patterns of nuclear CUB, but group in completely different clusters based on mitochondrial CUB. This suggests that different pressures may govern CUB in mitochondrial genomes than in their nuclear counterparts.

Patterns of codon usage bias reflect whether genes are mitochondrial or nuclear in origin

To examine signatures of codon usage between nuclear and mitochondrial genes, RSCU values for each gene in each available genome were calculated. A correspondence analysis (CA) was then performed to determine which codons drive observed differences in codon usage patterns (Figure 5). The CA plot (Figure 5A) shows a distinct clustering of the majority of the mitogenes away from nuclear genes. This demonstrates that codon usage signatures depend more on whether genes are mitochondrial versus nuclear as opposed to whether genes belong to the same species. The factor map of codon contributions (Figure 5B) revealed that the first dimension explains 15.6% of observed variation between genes in the final plot. The second dimension explains 7% of observed variance. Examining dimensional contributions by codon reveals that the GC content of the third position drives separation along dimensions. Position along the first

346 dimension (X-axis) is driven primarily by the usage of NNA versus NNC codons. The largest
347 contributions along the X-axis come from the usage of AUA (isoleucine) and CCC (proline).
348 RSCU values of greater than 1 indicate that a codon is overrepresented within a given
349 synonymous codon group whereas RSCU values less than 1 indicate underrepresentation. The
350 average RSCU of AUA and CCC are 1.6304 and 0.0449 in the mitochondria and 0.4525 and
351 1.0457 in the nucleus, respectively. Position along the second dimension (Y-axis) is driven
352 primarily by differences in the usage of NNU versus NNG codons. The largest contributions
353 along the Y-axis are from CCU (proline) versus GGG (glycine), ACG (threonine), and CCG
354 (proline) combined. The average RSCU of CCU, GGG, ACG, and CCG are 2.7140, 0.0558,
355 0.0204, and 0.0484 in the mitochondria and 1.072, 0.7240, 0.8274, and 0.8920 in the nucleus,
356 respectively.
357
358 A second CA was run using the RSCU values for each mitogene to determine which codons
359 drives observed interspecies differences in codon usage patterns in mitogenomes (Figure 6). The
360 CA plot shows distinct grouping based on gene identity as opposed to species of origin (Figure
361 6A). The factor map of codon contributions revealed that the first dimension explains 18.7% of
362 observed variation in the final CA plot, while the second-dimension accounts for 16.1% (Figure
363 6B). The *A. avenaceus atp8* gene is a clear outlier along both axes. The codons that contribute
364 the most to this are ACC, UCC and CCG which are used at a frequency of 4, 4, 1.33 respectively
365 in this gene. RSCU values of 4 indicate that only ACC (threonine) and UCC (serine)—none of
366 the other synonymous codons within their respective families—are used in this gene. This degree
367 of bias is expected given that threonine and serine occur only once and twice, respectively, in *A.*
368 *avenaceus atp8.*

Deviation of gene-level codon usage from neutral expectation varies based on whether genes are of nuclear or mitochondrial origin

To assess the role of mutational bias across all mitochondrial and nuclear genes, we examined the relationship between the ENc of each gene and its GC3 content by comparing observed ENc values to the expected relationship between ENC and GC3 content if codon usage was influenced by neutral mutational bias alone. We tested the fit to the neutral expectation of the complete dataset of all species' combined nuclear and mitochondrial gene datasets as well as all nuclear genes and all mitogenes separately by calculating the R^2 value. For all 20 species, combined nuclear and mitochondrial datasets yielded R^2 values greater than 0.5, suggesting that codon usage in these species can be partially explained by neutral mutational bias (Supplemental File S2). Furthermore, patterns of deviation from the neutral expectation were highly similar between species (Supplemental File S2). However, when nuclear and mitochondrial genes were analyzed separately, nuclear genes had an R^2 value of 0.598, whereas mitochondrial genes had an R^2 value of 0.211 (Figure 7). This suggests that, although codon usage in nuclear genomes can be partially explained by neutral mutational bias, mutational bias does not fully account for the codon bias in mitochondrial genomes.

Codon usage in nuclear genomes, but not mitogenomes, is under translational selection

To test if translational selection could account for the observed deviations of CUB from the neutral expectation, we calculated the S-values for each mitochondrial and nuclear genome. Of the 20 *Flavi* species tested, mitogenome S-values ranged from -0.103 to 0.392 with a median value of 0.162 and mean value of 0.137 (Figure 8A). However, no species' mitogenomes had S-

values that were found to be significant in the permutation test. In contrast, nuclear genome S-values ranged from 0.269 to 0.502, with a median value of 0.432 and a mean value of 0.427 (Figure 8B). The S-value of *A. novoparasiticus* (S = 0.269) was calculated using the package tAI.R (https://github.com/mariodosreis/tai/blob/master/R/tAI.R). This was done as the original calculation with stAI calc created issues with file merging. All nuclear S-values were found to be significant in the permutation test, suggesting that *Flavi* nuclear genomes are under moderate levels of translational selection.

Discussion

In this study, we compared the evolution of mitochondrial and nuclear genomes within *Aspergillus* section *Flavi*. We assembled and annotated the mitogenomes of 18 *Flavi* species and reannotated two previously assembled reference mitogenomes. We then used phylogenetic analyses to compare phylogenies derived from nuclear versus mitochondrial data. Finally, we examined the patterns of and forces underlying CUB in nuclear and mitochondrial genomes.

The newly assembled mitogenomes are comparable in gene content and size to previously published *Aspergillus* mitogenomes. At 29.10 kb to 39.27 kb, the range of *Aspergillus* section *Flavi* mitogenome length falls within the lower range of published fungal mitogenomes, which vary in size from 12.06 kb to 235.85 kb (Joardar et al. 2012; Zhang et al. 2020). GC content was consistent with low percentages observed in other *Aspergillus* and related fungal species (Machida et al. 2005; Sato et al. 2011; Joardar et al. 2012; Zhao et al. 2012; Yan et al. 2016; Park et al. 2019a; Park et al. 2019b; Park et al. 2020; Hugaboom et al. 2021). The mitogenomic content and gene order were highly conserved in the 20 *Flavi* species analyzed, and all mitogenomes examined contained 14 core mitochondrial genes. As in previous studies, these

core genes showed high levels of sequence similarity and conservation of gene order (Joardar et al. 2012; Hugaboom et al. 2021). Fungal mitogenomes are also known to contain accessory genes in addition to the core set of 14. The presence of the two accessory genes—an intron encoded LAGLIDADG endonuclease and the ribosomal protein S3—in most of the species analyzed is also consistent with existing *Flavi* annotations (Joardar et al. 2012; Hugaboom et al. 2021). The order of these accessory genes was also highly conserved. Of note, the mitogenomes contained their own set of 26 tRNAs separate from the nuclear-encoded set of tRNAs. In analyses of CUB, the mitochondrial tRNAs were used to determine if mitochondrial CUB patterns had been optimized to the mitochondrial tRNA pool.

In comparing the topologies and evolutionary rates predicted by phylogenies derived from nuclear and mitochondrial data, we found that their inferred evolutionary histories were similar. The high degree of congruence in the two phylogenies suggests potential coevolution of mitochondrial and nuclear genes. The minor disagreements between the two phylogenies may be explained by phenomena that occur uniquely in mitochondria. For example, fungal mitochondria are uniparentally inherited (Horn 2016 ;Santamaria et al. 2009). Additionally, fungal species can undergo interspecific hybridization (Giordano et al. 2018). In this process, the mitochondria of one species may be inherited by the other. Moreover, mitochondrial recombination events with repeated backcrossing can lead to introgression (Giordano et al. 2018). Interspecific introgression and recombination occur in fungal nuclei as well. Thus, the two phylogenies may differ due to disparities in interspecific introgression and/or recombination occurring in the mitochondria or nuclei of section *Flavi.* Alternatively, however, topological differences could

arise due to sampling error, as mitochondrial genes contain few sites relative to nuclear genes, for example.

Cluster analyses based on net RSCU values demonstrated that interspecies similarities in patterns of CUB differ between nuclear and mitochondrial genomes. Despite some parallels in predicted grouping—for example, the grouping of *A. caelatus* and *A. pseudocaelatus* in both dendrograms —mitochondrial and nuclear cluster analyses displayed groupings largely inconsistent with each other. It is important to note that all the species included in this study have similar mean values of codon usage metrics (ENc, GC content, and GC3s) within the nuclear and mitochondrial genome. Thus, well-resolved interspecies relationships are unlikely to be based on codon usage indices alone. Alternatively, the observed incongruence may reflect different pressures governing CUB in mitochondrial compared to nuclear genomes.

Examination of codon usage patterns using correspondence analyses showed that differential usage of certain codons drives observable differences in signatures of CUB between mitochondrial and nuclear genes and between gene type in mitogenes. Differential usage of specific codons between nuclear and mitochondrial genomes appears to rely heavily on the GC content of the third position of synonymous codons. This pattern aligns with overall GC content of the genomes. For example, the use of the codon AUA contributes to the placement of the mitogenomes in quadrant II of the final correspondence analysis plot, where mitogenes tend to cluster, while the use of AUC contributes to the placement of the nuclear genes in quadrants I and IV. The average RSCU values of AUA and AUC are 1.6304 and 0.1782 in the mitochondria and 0.4525 and 1.4760 in the nucleus, respectively. Both of these codons code for isoleucine, yet

461 mitogenes are enriched for the AUA codon and nuclear genes for the AUC codon, as would be
462 expected based on the differences in GC content between the two genomes. The separation of
463 mitogenes is also dependent on the GC content of the third position. Figure 6 shows that, while
464 the use of most codons is similar amongst all mitogenes, the occurrence of a rare G- or C-ending
465 codon drives separation based on CUB patterns. This is especially clear in the case of the
466 outlying *A. avenaceus atp8* gene in Figure 6A, which is driven by the higher use of codons ACC,
467 UCC, and CCG. Despite a high degree of sequence conservation with the other 19 *atp8*
468 nucleotide sequences (Supplementary File S3), the change in a small number of nucleotides at
469 third codon positions results in a large visible separation in the correspondence analysis plot
470 (Figure 6A). This effect is amplified due to the short, highly conserved nature of the *atp8* gene
471 sequences. Overall, we found that gene-level RSCU values allow for observable differences in
472 CUB pattern based on the organelle of genomic origin and mitogene identity.
473
474 We also sought to determine the relative importance of neutral processes and natural selection on
475 shaping CUB in mitochondrial and nuclear genomes. Based on ENc-GC3 plots, most mitogenes
476 fell at least 20% from the neutral expectation, while most nuclear genomes fell within 10% of the
477 neutral expectation. These results reinforce previous findings that CUB varies at the gene-level
478 within a species (Sharp et al. 1988; L et al. 2004; LaBella et al. 2019). Of note, studies have
479 shown that greater divergence from the neutral expectation is moderately associated with
480 increased expression (Tsankov et al. 2010). Future avenues may examine the association
481 between the large residuals from the neutral expectation and expression levels of mitogenes.
482

The moderate to poor fit to the neutral expectation for nuclear and mitochondrial genes, respectively, suggests that mutational bias alone cannot account for the observed patterns in codon usage bias. By using the S-test to test for the influence of translational selection, we found that gene-level codon usage in mitochondrial genomes could not be significantly distinguished from neutral mutational bias—including both selectively neutral changes and purifying selection—in section *Flavi*, while translational selection acts moderately on codon usage bias in nuclear genomes. The lack of significant translational selection on mitogenomes is unsurprising, given their extreme GC bias and small size. This may be a manifestation of mtDNA evolving clonally with limited ability to recombine; thus, CUB is more likely to reflect mutation bias and drift rather than selection. The faster evolutionary rates of mitogenomes may lead to genetic drift playing a larger role in shaping CUB than in corresponding nuclear genomes. Additionally, when genome size is small, it is hypothesized that low tRNA redundancy limits the ability of selection to act on CUB (dos Reis et al. 2004). Of note, S-value calculation for mitogenes was limited to a dataset of 16 genes. Visual inspection of the data used to determine the S-values suggests a general positive correlation between selective pressure and codon usage – which would suggest translational selection on codon usage – that is obscured by a couple outlier genes (Figure 8A). This observation in combination with the highly variable codon usage between mitochondrial genes suggests that the balance between selective and neutral forces on mitochondrial codon usage may vary greatly between mitogenes. Finally, the final S-value calculations for mitogenomes were based solely on the mitochondrial tRNA counts derived from genomic sequences and not experimental tRNA abundances. In fact, our analysis suggests that additional tRNA dynamics, such as modification or importation, may be at work in *Aspergillus* mitochondria.

Computational analysis of codon usage and tRNA composition in *Aspergillus* mitogenomes suggests that there is a significant gap in our knowledge of tRNA dynamics within these organelles. It is known that mitochondria can employ diverse strategies to obtain a complete and functional set of tRNAs; some organisms such as the fungus *Saccharomyces cerevisiae* encode a complete set of tRNA genes within their mitochondria (Salinas-Giegé et al. 2015) while others require the importation of nuclear tRNAs into the mitochondria (Alfonzo and Söll 2009). Our analysis demonstrates that , 15 codons cannot be decoded by the mitochondrial tRNAome without invoking liberal wobble base pairing, mitochondrial tRNA modification, or import of nuclear tRNAs(Supplementary File S4). This suggests that tRNA import or modification may be occurring in *Aspergillus* mitochondria. Additionally, mitochondrial codon usage is not consistently biased towards codons matching the mitochondrial tRNAome. For example, the codon GCA (alanine), which can be decoded by a mitochondrial tRNA, has an average RSCU value of 1.4222 in mitogenomes, whereas the codon GCU (also alanine) has an average RSCU value of 2.3423 even though the mitochondrial tRNAome is unable to decode this codon. The preference for GCU codons suggests the importation or modification of a tRNA capable of decoding this codon. Finally, the *Aspergillus* mitochondrial tRNAs fit the wobble versatility hypothesis for each codon family, with the exception of CGN (arginine), UGR (tryptophan), and AUR (methionine), a finding that is consistent with previous investigation of the wobble nucleotide position in fungal mitogenomes (Supplemental File S5) (Carullo and Xia 2008). That is, the anticodons of the mitochondrial tRNAome have nucleotides at the wobble site that maximize versatility in wobble base pairing as opposed to maximizing Watson-Crick base pairing with the most frequently used codon within each synonymous codon family. Improving

our understanding of *Aspergillus* mitochondrial tRNA dynamics will not only allow us to better understand translational dynamics within the organelle but recent work has suggested that mitochondrial tRNAs may play a role in antifungal response (Colabardini et al. 2022).

Despite a limited understanding of tRNA dynamics within *Aspergillus* mitochondria our results are consistent with the limited role of translation selection in shaping general patterns of mitochondrial codon usage in other species including budding yeasts, plants, and animals (Kamatani and Yamamoto 2007; Jia and Higgs 2008; Zhou and Li 2009). As with previous work, we also noted a few specific codons (proline codons) and genes (*atp8*) with increased biases that may be related to factors such as wobble-decoding or tRNA abundance.

In summary, analysis of mitochondrial and nuclear genome data from *Aspergillus* section *Flavi* revealed that both genomes are largely phylogenetically congruent and that the pattern and evolutionary forces shaping CUB differ between the mitochondrial and nuclear genomes. These evolutionary analyses, coupled with the generation of mitogenome assemblies for 18 section *Flavi* species, contribute to our understanding of genome evolution in the genus *Aspergillus.*

Data Availability Statement

The newly assembled *Aspergillus* section *Flavi* mitogenomes from this study are available in GenBank under accession numbers ON833077, ON833078, ON833079, ON833081, ON833082, ON833083, ON833084, ON833085, ON833086, ON833087, ON833088, ON833089, ON833090, ON833091, ON833092, ON833093, and ON833094. Reannotations for previously assembled mitogenomes are available through figshare (10.6084/m9.figshare.20412186). The

SRA accession numbers for whole genome sequencing data used for mitogenome assembly are provided in Table 1. For previously assembled mitogenomes, the NCBI reference sequence GenBank accession numbers are provided in place of SRA accession numbers. Additional data, including supplementary data, are available through figshare (https://doi.org/10.6084/m9.figshare.20412186).

Acknowledgements

We thank members of the Rokas Laboratory for their support and feedback. This work was performed in part using resources contained within the Advanced Computing Center for research and Education at Vanderbilt University in Nashville, TN.

Funding

M.H. was partially supported by a Vanderbilt Data Science Institute Summer Research Program (DSI – SRP) fellowship and E.A.H. by the National Institutes of Health/National Eye Institute (F31 EY033235). Research in A.R.'s lab is supported by grants from the National Science Foundation (DEB-2110404), the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R56 AI146096 and R01 AI153356), and the Burroughs Wellcome Fund.

Conflict of Interest Statement

A. R. is a scientific consultant for LifeMine Therapeutics, Inc. A. R. is also an Associate Editor for the *G3* journal.

Supplementary Materials

Available through figshare (10.6084/m9.figshare.20412186).

References

Alfonzo JD, Söll D. 2009. Mitochondrial tRNA import--the challenge to understand has just begun. Biol Chem. 390(8):717–722. doi:10.1515/BC.2009.101. https://pubmed.ncbi.nlm.nih.gov/19558325.

Basse CW. 2010. Mitochondrial inheritance in fungi. Curr Opin Microbiol. 13(6):712–9. doi:10.1016/j.mib.2010.09.003.

Brown TA, Waring RB, Scazzocchio C, Davies RW. 1985. The *Aspergillus nidulans* mitochondrial genome. Curr Genet. 9(2):113–117. doi:10.1007/BF00436957. https://doi.org/10.1007/BF00436957.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics. 129(3):897–907. doi:10.1093/genetics/129.3.897. https://pubmed.ncbi.nlm.nih.gov/1752426.

Burger G, Gray MW, Franz Lang B. 2003. Mitochondrial genomes: anything goes. Trends in Genetics. 19(12):709–716. doi:https://doi.org/10.1016/j.tig.2003.10.012. https://www.sciencedirect.com/science/article/pii/S0168952503003044.

Calderone R, Li D, Traven A. 2015. System-level impact of mitochondria on fungal virulence: to metabolism and beyond. FEMS Yeast Research. 15(4). doi:10.1093/femsyr/fov027. https://doi.org/10.1093/femsyr/fov027.

Chan PP, Lin BY, Mak AJ, Lowe TM. 2021. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Research. 49(16):9077–9096. doi:10.1093/nar/gkab688. https://doi.org/10.1093/nar/gkab688.

Chatre L, Ricchetti M. 2014. Are mitochondria the Achilles' heel of the Kingdom Fungi? Current Opinion in Microbiology. 20:49–54. doi:https://doi.org/10.1016/j.mib.2014.05.001. https://www.sciencedirect.com/science/article/pii/S1369527414000484.

Chen C, Li Q, Fu R, Wang J, Deng G, Chen X, Lu D. 2021. Comparative mitochondrial genome analysis reveals intron dynamics and gene rearrangements in two *Trametes* species. Scientific Reports. 11(1):2569. doi:10.1038/s41598-021-82040-7. https://doi.org/10.1038/s41598-021-82040-7.

Chevance FF v, le Guyon S, Hughes KT. 2014. The Effects of Codon Context on In Vivo Translation Speed. PLOS Genetics. 10(6):e1004392-. https://doi.org/10.1371/journal.pgen.1004392.

Coenen A, Croft JH, Slakhorst M, Debets F, Hoekstra R. 1996. Mitochondrial inheritance in *Aspergillus nidulans*. Genet Res. 67(2):93–100. doi:DOI: 10.1017/S0016672300033553. https://www.cambridge.org/core/article/mitochondrial-inheritance-in-aspergillus-nidulans/A5C8907D7AA836E947FE2F5B59BE1B71.

Colabardini AC, van Rhijn N, LaBella AL, Valero C, Wang F, Dineen L, Bromley MJ, Rokas A, Goldman GH. 2022 Jan 1. *Aspergillus fumigatus* FhdA transcription factor is important for mitochondrial activity and codon usage regulation during the caspofungin paradoxical effect. bioRxiv.:2022.05.21.492902. doi:10.1101/2022.05.21.492902. http://biorxiv.org/content/early/2022/05/25/2022.05.21.492902.abstract.

Dolezal AL, Shu X, OBrian GR, Nielsen DM, Woloshuk CP, Boston RS, Payne GA. 2014. *Aspergillus flavus* infection induces transcriptional and physical changes in developing maize kernels. Front Microbiol. 5. https://www.frontiersin.org/articles/10.3389/fmicb.2014.00384.

Edler D, Klein J, Antonelli A, Silvestro D. 2021. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. Methods Ecol Evol. 12(2):373–377. doi:https://doi.org/10.1111/2041-210X.13512. https://doi.org/10.1111/2041-210X.13512.

625 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative
626 genomics. Genome Biology. 20(1):238. doi:10.1186/s13059-019-1832-y.
627 https://doi.org/10.1186/s13059-019-1832-y.
628 Frisvad JC, Møller LLH, Larsen TO, Kumar R, Arnau J. 2018. Safety of the fungal workhorses
629 of industrial biotechnology: update on the mycotoxin and secondary metabolite potential of
630 *Aspergillus niger, Aspergillus oryzae*, and *Trichoderma reesei*. Appl Microbiol Biotechnol.
631 102(22):9481–9515. doi:10.1007/s00253-018-9354-1.
632 https://pubmed.ncbi.nlm.nih.gov/30293194.
633 Giordano L, Sillo F, Garbelotto M, Gonthier P. 2018. Mitonuclear interactions may contribute to
634 fitness of fungal hybrids. Scientific Reports. 8(1):1706. doi:10.1038/s41598-018-19922-w.
635 https://doi.org/10.1038/s41598-018-19922-w.
636 Gourama H, Bullerman LB. 1995. *Aspergillus flavus* and *Aspergillus parasiticus*: Aflatoxigenic
637 Fungi of Concern in Foods and Feeds†: A Review. Journal of Food Protection. 58(12):1395–
638 1404. doi:10.4315/0362-028X-58.12.1395. https://doi.org/10.4315/0362-028X-58.12.1395.
639 Gouy M, Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic
640 Acids Res. 10(22):7055–7074. doi:10.1093/nar/10.22.7055.
641 https://pubmed.ncbi.nlm.nih.gov/6760125.
642 Gray MW, Burger G, Lang BF. 1999. Mitochondrial Evolution. Science (1979). 283(5407):1476.
643 doi:10.1126/science.283.5407.1476.
644 http://science.sciencemag.org/content/283/5407/1476.abstract.
645 Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1:
646 expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res.
647 47(W1):W59–W64. doi:10.1093/nar/gkz238. https://pubmed.ncbi.nlm.nih.gov/30949694.

Grigoriev I v, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, et al. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research. 42(D1):D699–D704. doi:10.1093/nar/gkt1183. https://doi.org/10.1093/nar/gkt1183.

Gustafsson C, Govindarajan S, Minshull J. 2004. Codon bias and heterologous protein expression. Trends in Biotechnology. 22(7):346–353. doi:https://doi.org/10.1016/j.tibtech.2004.04.006. https://www.sciencedirect.com/science/article/pii/S0167779904001118.

Hanson G, Coller J. 2018. Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol. 19(1):20–30. doi:10.1038/nrm.2017.91. https://pubmed.ncbi.nlm.nih.gov/29018283.

Hedayati MT, Pasqualotto AC, Warn PA, Bowyer P, Denning DW. 2007. *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. Microbiology (N Y). 153(6):1677–1692. doi:10.1099/mic.0.2007/007641-0. [accessed 2021 May 2]. https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.2007/007641-0.

Hiraoka Y, Kawamata K, Haraguchi T, Chikashige Y. 2009. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. Genes to Cells. 14(4):499–509. doi:https://doi.org/10.1111/j.1365-2443.2009.01284.x. https://doi.org/10.1111/j.1365-2443.2009.01284.x.

Hoffmeister D, Keller NP. 2007. Natural products of filamentous fungi: enzymes, genes, and their regulation. Nat Prod Rep. 24(2):393–416. doi:10.1039/B603084J. http://dx.doi.org/10.1039/B603084J.

Homa M, Manikandan P, Szekeres A, Kiss N, Kocsubé S, Kredics L, Alshehri B, Dukhyil AA bin, Revathi R, Narendran V, et al. 2019. Characterization of *Aspergillus tamarii* Strains From Human Keratomycoses: Molecular Identification, Antifungal Susceptibility Patterns and Cyclopiazonic Acid Producing Abilities. Front Microbiol. 10:2249. doi:10.3389/fmicb.2019.02249. https://pubmed.ncbi.nlm.nih.gov/31649626.

Hatmaker EA, Rangel-Grimaldo M, Raja HA, Pourhadi H, Knowles SL, Fuller K, Adams EM, Lightfoot JD, Bastos RW, Goldman GH, et al. 2022 Jan 1. Genomic and phenotypic trait variation of the opportunistic human pathogen *Aspergillus flavus* and its non-pathogenic close relatives. bioRxiv.:2022.07.12.499845. doi:10.1101/2022.07.12.499845. http://biorxiv.org/content/early/2022/07/13/2022.07.12.499845.abstract.

Horn BW, Gell RM, Singh R, Sorensen RB, Carbone I. 2016. Sexual Reproduction in Aspergillus flavus Sclerotia: Acquisition of Novel Alleles from Soil Populations and Uniparental Mitochondrial Inheritance. PLoS One. 11(1):e0146169. doi:10.1371/journal.pone.0146169.

Hugaboom M, Beck ML, Carrubba KH, Chennupati DV, Gupta A, Liu Q, Reddy MK, Roque F, Hatmaker EA. 2021. Complete Mitochondrial Genome Sequences of Nine *Aspergillus flavus* Strains. Microbiol Resour Announc. 10(45):e0097121–e0097121. doi:10.1128/MRA.00971-21. https://pubmed.ncbi.nlm.nih.gov/34761953.

Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. Journal of Molecular Biology. 151(3):389–409. doi:https://doi.org/10.1016/0022-2836(81)90003-6. https://www.sciencedirect.com/science/article/pii/0022283681900036.

692 Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms.
693 Molecular Biology and Evolution. 2(1):13–34. doi:10.1093/oxfordjournals.molbev.a040335.
694 https://doi.org/10.1093/oxfordjournals.molbev.a040335.
695 Jia W, Higgs PG. 2008. Codon Usage in Mitochondrial Genomes: Distinguishing Context-
696 Dependent Mutation from Translational Selection. Molecular Biology and Evolution. 25(2):339–
697 351. doi:10.1093/molbev/msm259. https://doi.org/10.1093/molbev/msm259.
698 Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a
699 fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology.
700 21(1):241. doi:10.1186/s13059-020-02154-5. https://doi.org/10.1186/s13059-020-02154-5.
701 Joardar V, Abrams NF, Hostetler J, Paukstelis PJ, Pakala S, Pakala SB, Zafar N, Abolude OO,
702 Payne G, Andrianopoulos A, et al. 2012. Sequencing of mitochondrial genomes of nine
703 *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main
704 sources of genome size variability. BMC Genomics. 13(1):698. doi:10.1186/1471-2164-13-698.
705 https://doi.org/10.1186/1471-2164-13-698.
706 Juhász Á, Pfeiffer I, Keszthelyi A, Kucsera J, Vágvölgyi C, Hamari Z. 2008. Comparative
707 analysis of the complete mitochondrial genomes of *Aspergillus niger* mtDNA type 1a and
708 *Aspergillus tubingensis* mtDNA type 2b. FEMS Microbiology Letters. 281(1):51–57.
709 doi:10.1111/j.1574-6968.2008.01077.x. https://doi.org/10.1111/j.1574-6968.2008.01077.x.
710 Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment,
711 interactive sequence choice and visualization. Briefings in Bioinformatics. 20(4):1160–1166.
712 doi:10.1093/bib/bbx108. https://doi.org/10.1093/bib/bbx108.
713 Kamatani T, Yamamoto T. 2007. Comparison of codon usage and tRNAs in mitochondrial
714 genomes of *Candida* species. Biosystems. 90(2):362–370.

715 doi:https://doi.org/10.1016/j.biosystems.2006.09.039.
716 https://www.sciencedirect.com/science/article/pii/S0303264706002048.
717 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
718 Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop
719 software platform for the organization and analysis of sequence data. Bioinformatics.
720 28(12):1647–1649. doi:10.1093/bioinformatics/bts199.
721 https://pubmed.ncbi.nlm.nih.gov/22543367.
722 Kjærbølling I, Vesth T, Frisvad JC, Nybo JL, Theobald S, Kildgaard S, Petersen TI, Kuo A, Sato
723 A, Lyhne EK, et al. 2020. A comparative genomics study of 23 *Aspergillus* species from section
724 *Flavi*. Nature Communications. 11(1):1106. doi:10.1038/s41467-019-14051-y.
725 https://doi.org/10.1038/s41467-019-14051-y.
726 L CS, William L, K HA, Lucy S, H MH. 2004. Codon usage between genomes is constrained by
727 genome-wide mutational processes. Proceedings of the National Academy of Sciences.
728 101(10):3480–3485. doi:10.1073/pnas.0307827100. https://doi.org/10.1073/pnas.0307827100.
729 LaBella AL, Opulente DA, Steenwyk JL, Hittinger CT, Rokas A. 2019. Variation and selection
730 on codon usage bias across an entire subphylum. PLoS Genet. 15(7):e1008304–e1008304.
731 doi:10.1371/journal.pgen.1008304. https://pubmed.ncbi.nlm.nih.gov/31365533.
732 LaBella AL, Opulente DA, Steenwyk JL, Hittinger CT, Rokas A. 2021. Signatures of optimal
733 codon usage in metabolic genes inform budding yeast ecology. PLOS Biology. 19(4):e3001185-.
734 https://doi.org/10.1371/journal.pbio.3001185.
735 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods.
736 9(4):357–359. doi:10.1038/nmeth.1923. https://doi.org/10.1038/nmeth.1923.

Lavín JL, Oguiza JA, Ramírez L, Pisabarro AG. 2008. Comparative genomics of the oxidative phosphorylation system in fungi. Fungal Genetics and Biology. 45(9):1248–1256. doi:https://doi.org/10.1016/j.fgb.2008.06.005. https://www.sciencedirect.com/science/article/pii/S1087184508001084.

Leinonen R, Sugawara H, Shumway M, Collaboration INSD. 2011. The sequence read archive. Nucleic Acids Res. 39(Database issue):D19–D21. doi:10.1093/nar/gkq1019. https://pubmed.ncbi.nlm.nih.gov/21062823.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352. https://pubmed.ncbi.nlm.nih.gov/19505943.

Li Q, Wang Q, Jin X, Chen Z, Xiong C, Li P, Zhao J, Huang W. 2019. Characterization and comparison of the mitochondrial genomes from two *Lyophyllum* fungal species and insights into phylogeny of Agaricomycetes. International Journal of Biological Macromolecules. 121:364–372. doi:https://doi.org/10.1016/j.ijbiomac.2018.10.037. https://www.sciencedirect.com/science/article/pii/S0141813018332653.

Li Q, Xiang D, Wan Y, Wu Q, Wu X, Ma C, Song Y, Zhao G, Huang W. 2019. The complete mitochondrial genomes of five important medicinal *Ganoderma* species: Features, evolution, and phylogeny. International Journal of Biological Macromolecules. 139:397–408. doi:https://doi.org/10.1016/j.ijbiomac.2019.08.003. https://www.sciencedirect.com/science/article/pii/S0141813019326443.

758 Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto K-I, Arima T, Akita O,
759 Kashiwagi Y, et al. 2005. Genome sequencing and analysis of *Aspergillus oryzae*. Nature.
760 438(7071):1157–1161. doi:10.1038/nature04300. https://doi.org/10.1038/nature04300.
761 Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable
762 and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics.
763 35(21):4453–4455. doi:10.1093/bioinformatics/btz305.
764 https://doi.org/10.1093/bioinformatics/btz305.
765 Machida M, Yamada O, Gomi K. 2008. Genomics of *Aspergillus oryzae*: learning from the
766 history of Koji mold and exploration of its future. DNA Res. 15(4):173–183.
767 doi:10.1093/dnares/dsn020. https://pubmed.ncbi.nlm.nih.gov/18820080.
768 Martins VP, Dinamarco TM, Soriani FM, Tudella VG, Oliveira SC, Goldman GH, Curti C,
769 Uyemura SA. 2011. Involvement of an alternative oxidase in oxidative stress and mycelium-to-
770 yeast differentiation in *Paracoccidioides brasiliensis*. Eukaryot Cell. 10(2):237–248.
771 doi:10.1128/EC.00194-10. https://pubmed.ncbi.nlm.nih.gov/21183691.
772 Mukhopadhyay J, Hausner G. 2021. Organellar Introns in Fungi, Algae, and Plants. Cells.
773 10(8):2001. doi:10.3390/cells10082001.
774 Nakamura K, Pirtle RM, Pirtle IL, Takeishi K, Inouye M. 1980. Messenger ribonucleic acid of
775 the lipoprotein of the *Escherichia coli* outer membrane. II. The complete nucleotide sequence.
776 Journal of Biological Chemistry. 255(1):210–216. doi:https://doi.org/10.1016/S0021-
777 9258(19)86285-3. https://www.sciencedirect.com/science/article/pii/S0021925819862853.
778 Park J, Kwon W, Zhu B, Mageswari A, Heo I-B, Kim J-H, Han K-H, Hong S-B. 2019a.
779 Complete mitochondrial genome sequence of an aflatoxin B and G producing fungus,

Aspergillus parasiticus. Mitochondrial DNA Part B. 4(1):947–948.
doi:10.1080/23802359.2018.1558126. https://doi.org/10.1080/23802359.2018.1558126.
Park J, Lee M-K, Yu J-H, Kim J-H, Han K-H. 2020. Complete mitochondrial genome sequence of Afla-Guard®, commercially available non-toxigenic *Aspergillus flavus*. Mitochondrial DNA Part B. 5(3):3572–3574. doi:10.1080/23802359.2020.1825129.
https://doi.org/10.1080/23802359.2020.1825129.
Payne BL, Alvarez-Ponce D. 2019. Codon Usage Differences among Genes Expressed in Different Tissues of Drosophila melanogaster. Genome Biology and Evolution. 11(4):1054–1065. doi:10.1093/gbe/evz051. https://doi.org/10.1093/gbe/evz051.
Post LE, Strycharz GD, Nomura M, Lewis H, Dennis PP. 1979. Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in *Escherichia coli*. Proc Natl Acad Sci U S A. 76(4):1697–1701. doi:10.1073/pnas.76.4.1697.
https://pubmed.ncbi.nlm.nih.gov/377281.
Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, et al. 2015. Codon Optimality Is a Major Determinant of mRNA Stability. Cell. 160(6):1111–1124. doi:https://doi.org/10.1016/j.cell.2015.02.029.
https://www.sciencedirect.com/science/article/pii/S0092867415001956.
dos Reis M, Savva R, Wernisch L. 2004. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. 32(17):5036–5044. doi:10.1093/nar/gkh834.
https://pubmed.ncbi.nlm.nih.gov/15448185.
Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. 2017. Variant Review with the Integrative Genomics Viewer. Cancer Res. 77(21):e31–e34. doi:10.1158/0008-5472.CAN-17-0337. https://pubmed.ncbi.nlm.nih.gov/29092934.

803 Roymondal U, Das S, Sahoo S. 2009. Predicting Gene Expression Level from Relative Codon
804 Usage Bias: An Application to *Escherichia coli* Genome. DNA Research. 16(1):13–30.
805 doi:10.1093/dnares/dsn029. https://doi.org/10.1093/dnares/dsn029.
806 Sabi R, Volvovitch Daniel R, Tuller T. 2017. stAIcalc: tRNA adaptation index calculator based
807 on species-specific weights. Bioinformatics. 33(4):589–591. doi:10.1093/bioinformatics/btw647.
808 https://doi.org/10.1093/bioinformatics/btw647.
809 Sahoo S, Das SS, Rakshit R. 2019. Codon usage pattern and predicted gene expression in
810 *Arabidopsis thaliana*. Gene. 721:100012. doi:https://doi.org/10.1016/j.gene.2019.100012.
811 https://www.sciencedirect.com/science/article/pii/S2590158319300099.
812 Salinas-Giegé T, Giegé R, Giegé P. 2015. tRNA biology in mitochondria. Int J Mol Sci. 16(3).
813 Sanglard D, Ischer F, Bille J. 2001. Role of ATP-binding-cassette transporter genes in high-
814 frequency acquisition of resistance to azole antifungals in *Candida glabrata*. Antimicrob Agents
815 Chemother. 45(4):1174–1183. doi:10.1128/AAC.45.4.1174-1183.2001.
816 https://pubmed.ncbi.nlm.nih.gov/11257032.
817 Santamaria M, Vicario S, Pappadà G, Scioscia G, Scazzocchio C, Saccone C. 2009. Towards
818 barcode markers in Fungi: an intron map of *Ascomycota* mitochondria. BMC Bioinformatics.
819 10(6):S15. doi:10.1186/1471-2105-10-S6-S15. https://doi.org/10.1186/1471-2105-10-S6-S15.
820 Sato A, Oshima K, Noguchi H, Ogawa M, Takahashi T, Oguma T, Koyama Y, Itoh T, Hattori
821 M, Hanya Y. 2011. Draft genome sequencing and comparative analysis of *Aspergillus sojae*
822 NBRC4239. DNA Res. 18(3):165–176. doi:10.1093/dnares/dsr009.
823 https://pubmed.ncbi.nlm.nih.gov/21659486.
824 Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F. 1988. Codon usage patterns
825 in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe,*

826 *Drosophila melanogaster* and *Homo sapiens* ; a review of the considerable within-species
827 diversity. Nucleic Acids Research. 16(17):8207–8211. doi:10.1093/nar/16.17.8207.
828 https://doi.org/10.1093/nar/16.17.8207.
829 Sharp P M, Li WH. 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect
830 selection for “rare” codons. Nucleic Acids Res. 14(19):7737–7749. doi:10.1093/nar/14.19.7737.
831 https://pubmed.ncbi.nlm.nih.gov/3534792.
832 Sharp PM, Li W-H. 1986. An evolutionary perspective on synonymous codon usage in
833 unicellular organisms. Journal of Molecular Evolution. 24(1):28–38. doi:10.1007/BF02099948.
834 https://doi.org/10.1007/BF02099948.
835 Sharp PM, Stenico M, Peden JF, Lloyd AT. 1993. Codon usage: mutational bias, translational
836 selection, or both? Biochemical Society Transactions. 21(4):835–841. doi:10.1042/bst0210835.
837 https://doi.org/10.1042/bst0210835.
838 Shen X-X, Steenwyk JL, LaBella AL, Opulente DA, Zhou X, Kominek J, Li Y, Groenewald M,
839 Hittinger CT, Rokas A. 2020. Genome-scale phylogeny and contrasting modes of genome
840 evolution in the fungal phylum *Ascomycota*. Sci Adv. 6(45):eabd0079.
841 doi:10.1126/sciadv.abd0079. https://pubmed.ncbi.nlm.nih.gov/33148650.
842 Shields DC, Sharp PM. 1987. Synonymous codon usage in *Bacillus subtilis* reflects both
843 translational selection and mutational biases. Nucleic Acids Research. 15(19):8023–8040.
844 doi:10.1093/nar/15.19.8023. https://doi.org/10.1093/nar/15.19.8023.
845 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
846 large phylogenies. Bioinformatics. 30(9):1312–1313. doi:10.1093/bioinformatics/btu033.
847 https://pubmed.ncbi.nlm.nih.gov/24451623.

Steenwyk JL, Buida III TJ, Li Y, Shen X-X, Rokas A. 2020. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLOS Biology. 18(12):e3001007-. https://doi.org/10.1371/journal.pbio.3001007.

Stein A, Sia EA. 2017. Mitochondrial DNA repair and damage tolerance. Front Biosci (Landmark Ed). 22(5):920–943. doi:10.2741/4525.

Stoletzki N, Eyre-Walker A. 2007. Synonymous Codon Usage in *Escherichia coli*: Selection for Translational Accuracy. Molecular Biology and Evolution. 24(2):374–381. doi:10.1093/molbev/msl166. https://doi.org/10.1093/molbev/msl166.

Thomas LK, Dix DB, Thompson RC. 1988. Codon choice and gene expression: synonymous codons differ in their ability to direct aminoacylated-transfer RNA binding to ribosomes in vitro. Proc Natl Acad Sci U S A. 85(12):4242–4246. doi:10.1073/pnas.85.12.4242. https://pubmed.ncbi.nlm.nih.gov/3288988.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45(W1):W6–W11. doi:10.1093/nar/gkx391. https://pubmed.ncbi.nlm.nih.gov/28486635.

Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. 2010. The Role of Nucleosome Positioning in the Evolution of Gene Regulation. PLOS Biology. 8(7):e1000414-. https://doi.org/10.1371/journal.pbio.1000414.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 27(2):171–180. doi:https://doi.org/10.1111/j.1096-0031.2010.00329.x. https://doi.org/10.1111/j.1096-0031.2010.00329.x.

de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, Anderluh G, Asadollahi M, Askin M, Barry K, et al. 2017. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. Genome Biology. 18(1):28. doi:10.1186/s13059-017-1151-0. https://doi.org/10.1186/s13059-017-1151-0.

Wang X, Wang Y, Yao W, Shen J, Chen M, Gao M, Ren J, Li Q, Liu N. 2020. The 256 kb mitochondrial genome of *Clavaria fumosa* is the largest among phylum Basidiomycota and is rich in introns and intronic ORFs. IMA Fungus. 11(1):26. doi:10.1186/s43008-020-00047-7. https://doi.org/10.1186/s43008-020-00047-7.

Wei L, He J, Jia X, Qi Q, Liang Z, Zheng H, Ping Y, Liu S, Sun J. 2014. Analysis of codon usage bias of mitochondrial genome in *Bombyx moriand* its relation to evolution. BMC Evolutionary Biology. 14(1):262. doi:10.1186/s12862-014-0262-4. https://doi.org/10.1186/s12862-014-0262-4.

Wint R, Salamov A, Grigoriev I v. 2022. Kingdom-Wide Analysis of Fungal Protein-Coding and tRNA Genes Reveals Conserved Patterns of Adaptive Evolution. Mol Biol Evol. 39(2):msab372. doi:10.1093/molbev/msab372. https://doi.org/10.1093/molbev/msab372.Wright F. 1990. The 'effective number of codons' used in a gene. Gene. 87(1):23–29. doi:https://doi.org/10.1016/0378-1119(90)90491-9. https://www.sciencedirect.com/science/article/pii/0378111990904919.

Xia X. 1998. How optimized is the translational machinery in *Escherichia coli, Salmonella typhimurium* and *Saccharomyces cerevisiae*? Genetics. 149(1):37–44. doi:10.1093/genetics/149.1.37. https://pubmed.ncbi.nlm.nih.gov/9584084.

Xia X. 2017. DAMBE6: New Tools for Microbial Genomics, Phylogenetics, and Molecular Evolution. J Hered. 108(4):431–437. doi:10.1093/jhered/esx033. https://pubmed.ncbi.nlm.nih.gov/28379490.

Yan Z, Chen D, Shen Y, Ye B. 2016. The complete mitochondrial genome sequence of *Aspergillus flavus*. Mitochondrial DNA Part A. 27(4):2671–2672. doi:10.3109/19401736.2015.1022752. https://doi.org/10.3109/19401736.2015.1022752.

Zardoya R. 2020. Recent advances in understanding mitochondrial genome diversity. F1000Res. 9. doi:10.12688/f1000research.21490.1.

Zhang Y, Yang G, Fang M, Deng C, Zhang K-Q, Yu Z, Xu J. 2020. Comparative Analyses of Mitochondrial Genomes Provide Evolutionary Insights Into Nematode-Trapping Fungi. Frontiers in Microbiology. 11:617. https://www.frontiersin.org/article/10.3389/fmicb.2020.00617.

Zhao G, Yao Y, Qi W, Wang C, Hou L, Zeng B, Cao X. 2012. Draft genome sequence of *Aspergillus oryzae* strain 3.042. Eukaryot Cell. 11(9):1178. doi:10.1128/EC.00160-12. https://pubmed.ncbi.nlm.nih.gov/22933657.

Zhipeng Z, Yunkun D, Mian Z, Lin L, Chien-hung Y, Jingjing F, She C, Yi L. 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proceedings of the National Academy of Sciences. 113(41):E6117–E6125. doi:10.1073/pnas.1606724113. https://doi.org/10.1073/pnas.1606724113.

Zhou M, Li X. 2009. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. Mol Biol Rep. 36(8):2039–2046. doi:10.1007/s11033-008-9414-1. https://doi.org/10.1007/s11033-008-9414-1.

913 Zhou T, Weems M, Wilke CO. 2009. Translationally optimal codons associate with structurally
914 sensitive sites in proteins. Mol Biol Evol. 26(7):1571–1580. doi:10.1093/molbev/msp070.
915 https://pubmed.ncbi.nlm.nih.gov/19349643.
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933

Table 1: Summary of sources of sequencing data. Reference mitochondrial genomes were used for *Aspergillus sojae*, *A. oryzae*, and *A. niger*. Raw paired-end whole genome sequencing reads were used for the remaining species.

Species	SRA Accession/ NCBI Reference Sequence GenBank	Source
Aspergillus flavus	SRR18725159	Hatmaker et al. 2022
Aspergillus transmontanesis	SRR8398939	Kjærbølling et al. 2020
Aspergillus arachidicola	SRR8398876	Kjærbølling et al. 2020
Aspergillus nomiae	SRR19369914	Hatmaker et al. 2022
Aspergillus parasiticus	SRR8840397	Kjærbølling et al. 2020
Aspergillus sergii	SRR8840616	Kjærbølling et al. 2020
Aspergillus sojae	AP014506.1	Sato et al. 2011
Aspergillus oryzae	NC_008282.1	Machida et al. 2005
Aspergillus minisclerotigenes	SRR8398929	Kjærbølling et al. 2020
Aspergillus caelatus	SRR8840396	Kjærbølling et al. 2020
Aspergillus pseudocaelatus	SRR8840541	Kjærbølling et al. 2020
Aspergillus pseudotamarii	SRR8840579	Kjærbølling et al. 2020
Apsergillus tamarii	SRR8840604	Kjærbølling et al. 2020
Aspergillus pseudonomiae	SRR8840540	Kjærbølling et al. 2020
Aspergillus bertholletius	SRR8398880	Kjærbølling et al. 2020
Aspergillus alliaceus	SRR8396970	Kjærbølling et al. 2020

Aspergillus coremiiformis	SRR8398877	Kjærbølling et al. 2020
Aspergillus leporis	SRR8398928	Kjærbølling et al. 2020
Aspergillus avenaceus	SRR8839916	Kjærbølling et al. 2020
Aspergillus novoparasiticus	SRR8398934	Kjærbølling et al. 2020
Aspergillus niger	NC_007445.1	Juhasz et al. 2008

Table 2: Mitochondrial and nuclear *Aspergillus* genomes differ greatly in size, genomic content, and GC bias. The summary above includes ranges of values from 20 *Aspergillus* section *Flavi* species' mitochondrial and corresponding nuclear genomes.

	Genome length	rRNAs	tRNAs	CDS	GC%
Mitogenomes	29.100-39.269 Kbp	2	26	15-17	24.9-26.9%
Nuclear Genomes	30,1001-40,900 Kbp	Undetermined	228-272	9,078-14,216	43-48.8%

Figure Legends

Figure 1: The typical *Aspergillus* section *Flavi* mitogenome is a circular DNA molecule. Here, the circularized mitogenome of *Aspergillus flavus* NRRL 1957is visualized. The blocks around the outer circle indicate genes color coded by function. Each assembled section *Flavi* mitogenome shared a conserved set of 14 core mitochondrial genes, 2 rRNA genes, and 25-27 tRNA genes in the order pictured above. GC content (26.2% overall) is illustrated as the interior circle's gray region.

Figure 2: *Aspergillus* section *Flavi* mitogenomes demonstrate conserved gene content and order. Synteny plot of *Aspergillus* section *Flavi* mitochondrial protein-coding genes. Core mitochondrial genes are universally present in section *Flavi,* with conserved order. Each arrow represents a separate gene. Each line of arrows represents a different species. Connections between species illustrate nucleotide sequence conservation, with darker connections indicating higher similarity.

Figure 3: Phylogenies constructed from nuclear and mitochondrial data predict similar evolutionary relationships, with minor differences in inferred topology arising amongst more closely related species. A) Maximum likelihood phylogeny based on concatenation of 2,422 nuclear orthologs with bootstrap values from 1000 replicates. B) Maximum likelihood phylogeny based on concatenation of 14 core mitogene sequences with bootstrap values from 1000 replicates. Numbers above nodes indicate bootstrap values.

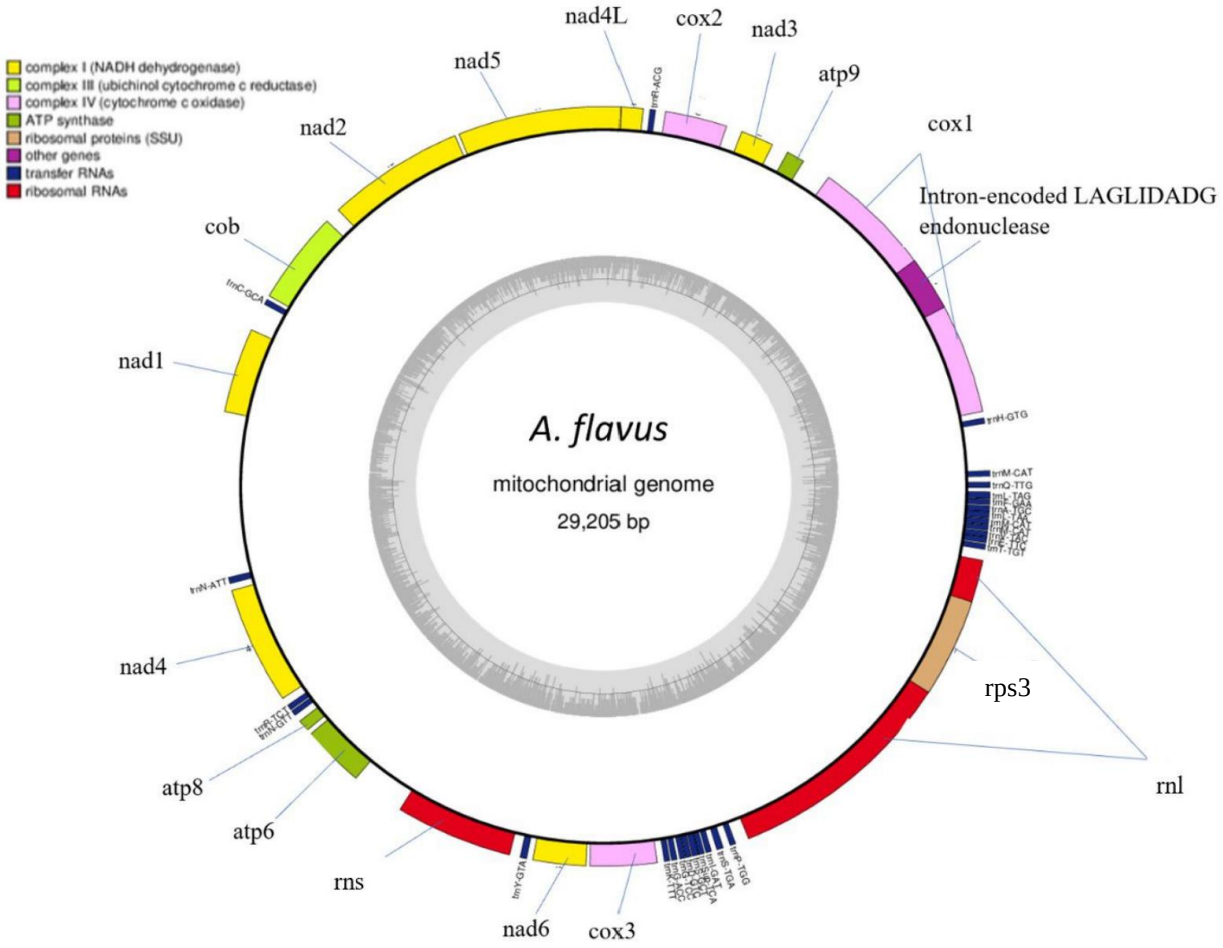
Figure 4: Hierarchical clustering analyses of relative synonymous codon usage (RSCU) values of mitochondrial and nuclear protein-coding regions demonstrate different species groupings based on mitochondrial and nuclear data. Dendrograms are colored/highlighted by species. A) Cluster analysis based on net RSCU values of nuclear protein-coding genes B) Cluster analysis based on net RSCU values of mitochondrial protein-coding genes.

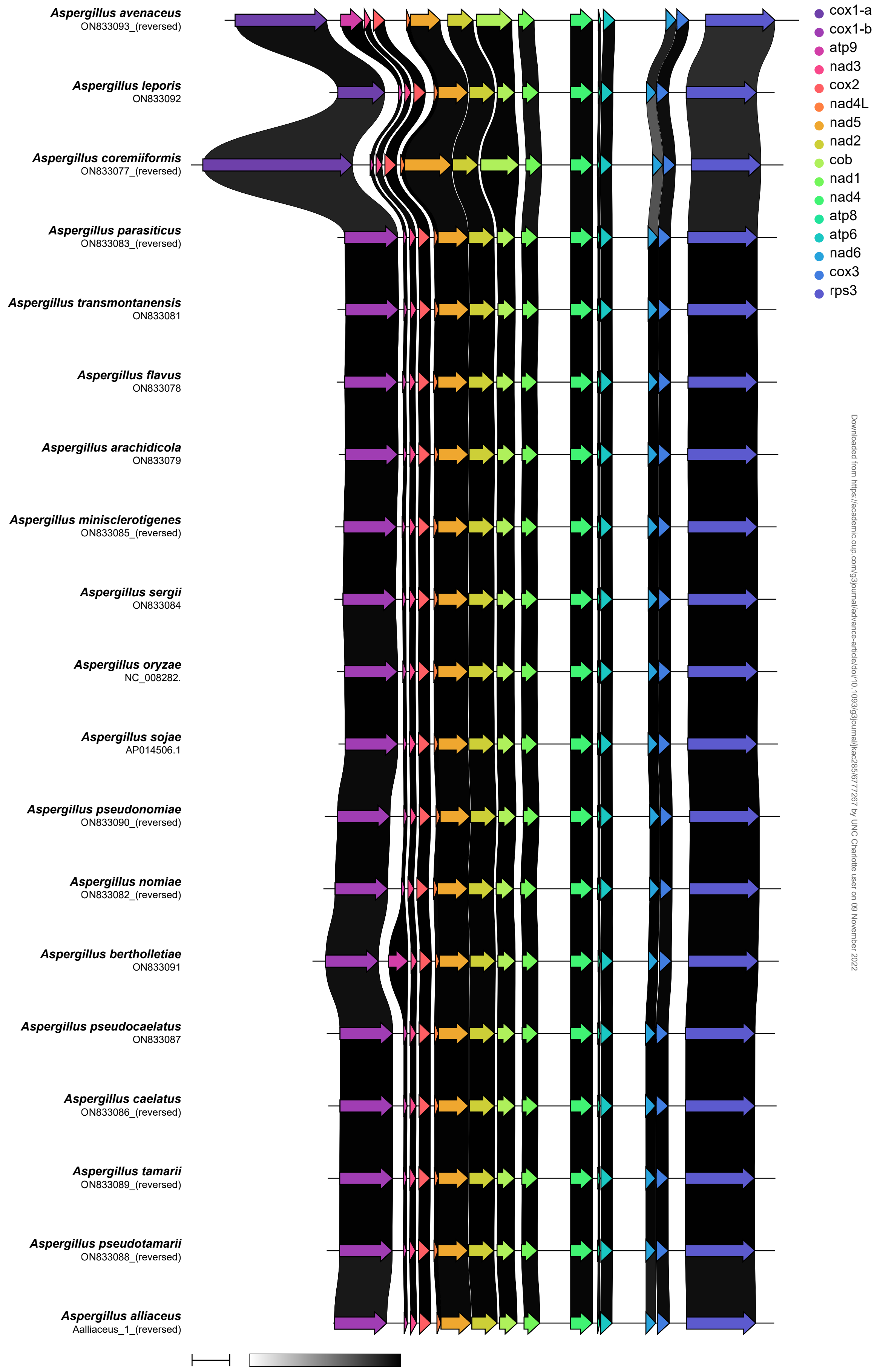
Figure 5: Correspondence Analysis based on relative synonymous codon usage values reveals that signatures of codon usage bias are more similar based on organelle of origin as opposed to species of origin. A) Correspondence analysis plot of all mitochondrial (yellow) and nuclear (purple) protein-coding genes for 20 *Aspergillus* section *Flavi* species. Each dot represents a gene. B) Factor map of codon contributions. Location of genes in correspondence analysis plot is driven largely by the GC content in the third position of synonymous codons used in the gene of interest.

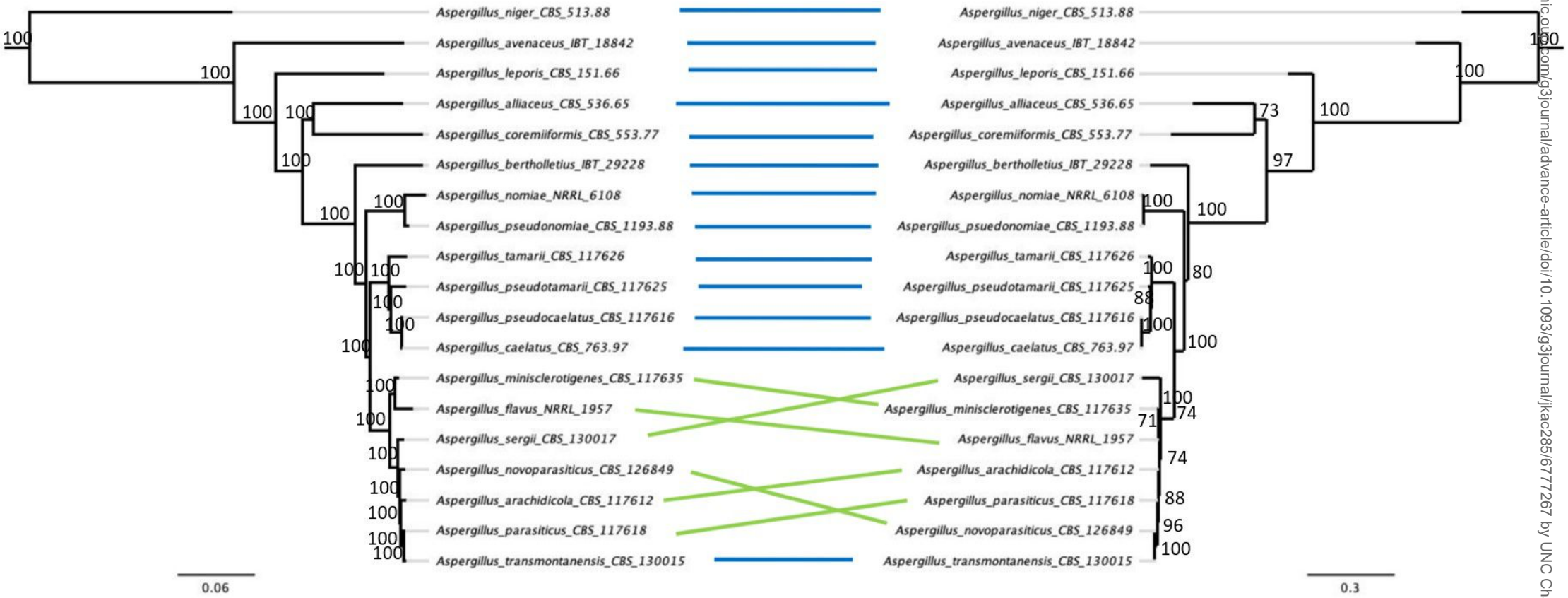
Figure 6: Correspondence Analysis based on relative synonymous codon usage values in mitogenes reveals that signatures of codon usage bias are more similar based on gene identity as opposed to species of origin. A) Correspondence analysis plot of all mitochondrial protein-coding genes for 20 *Aspergillus* section *Flavi* species. Labels correspond to gene identity B) Factor map of codon contributions

Figure 7: Most signatures of codon usage bias in mitochondrial and nuclear genes in *Aspergillus* section *Flavi* deviate from the expected codon usage bias under mutation pressure alone. A) ENc-GC3 plot for all nuclear protein-coding genes of 20 *Aspergillus* section

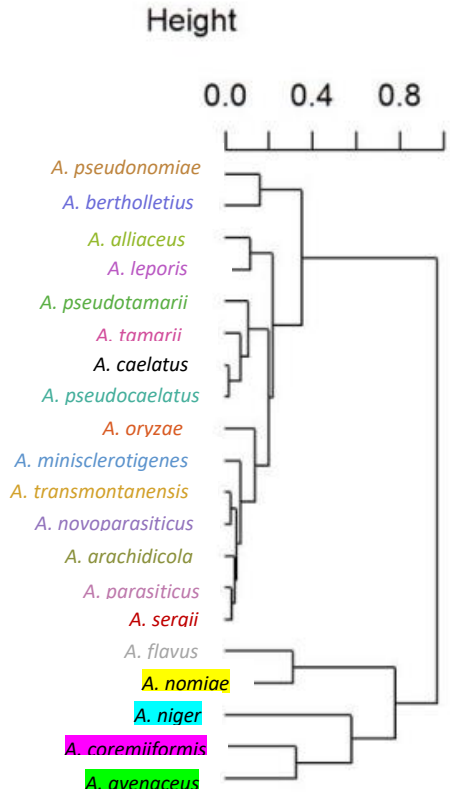
990 Flavi genes plotted against the predicted neutral distribution. R^2 value of 0.598 indicates
991 moderate fit to neutral expectation. B) ENc-GC3 plot for all protein-coding mitogenes. R^2 value
992 of 0.211 indicates poor fit to neutral expectation.
993
994 **Figure 8: Section *Flavi* mitogenomes are not under significant translational selection on**
995 **codon usage bias, but nuclear genomes display moderate translational selection.** Plots of
996 stAI against selective pressure for all protein-coding genes of *Aspergillus flavus* A) Mitogenes
997 only. Example of insignificant translational selection on S-test (S=0.191). B) Nuclear genes only.
998 Example of moderate translational selection on S-test (S=0.454).
999



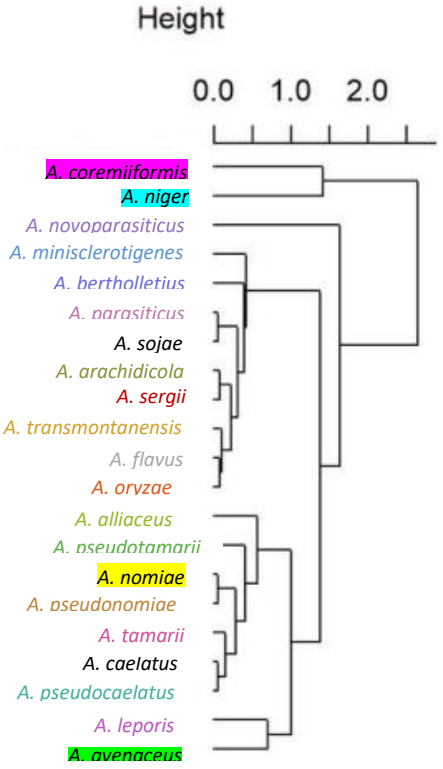




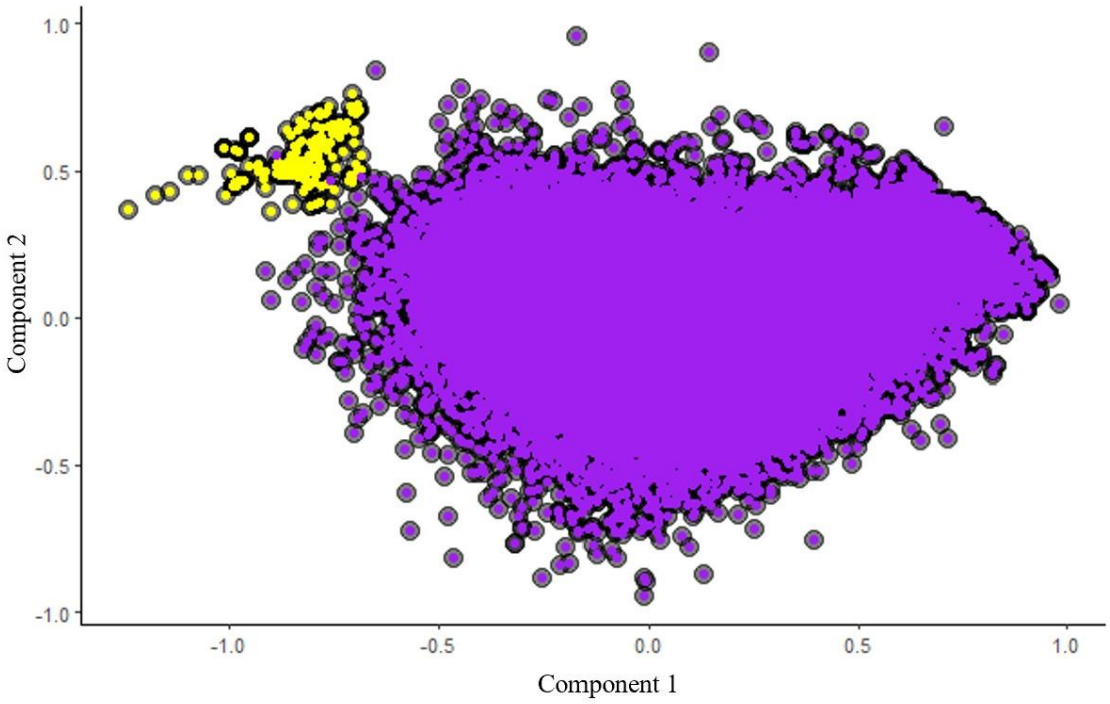
A. Nuclear Cluster Dendrogram



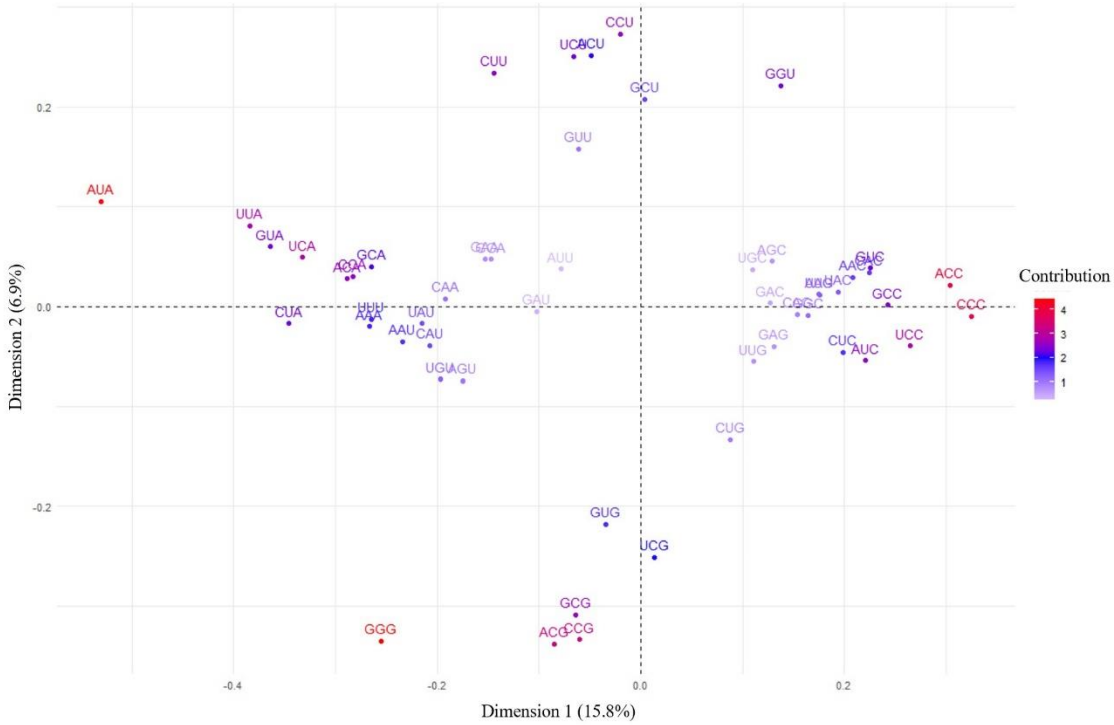
B. Mitochondrial Cluster Dendrogram



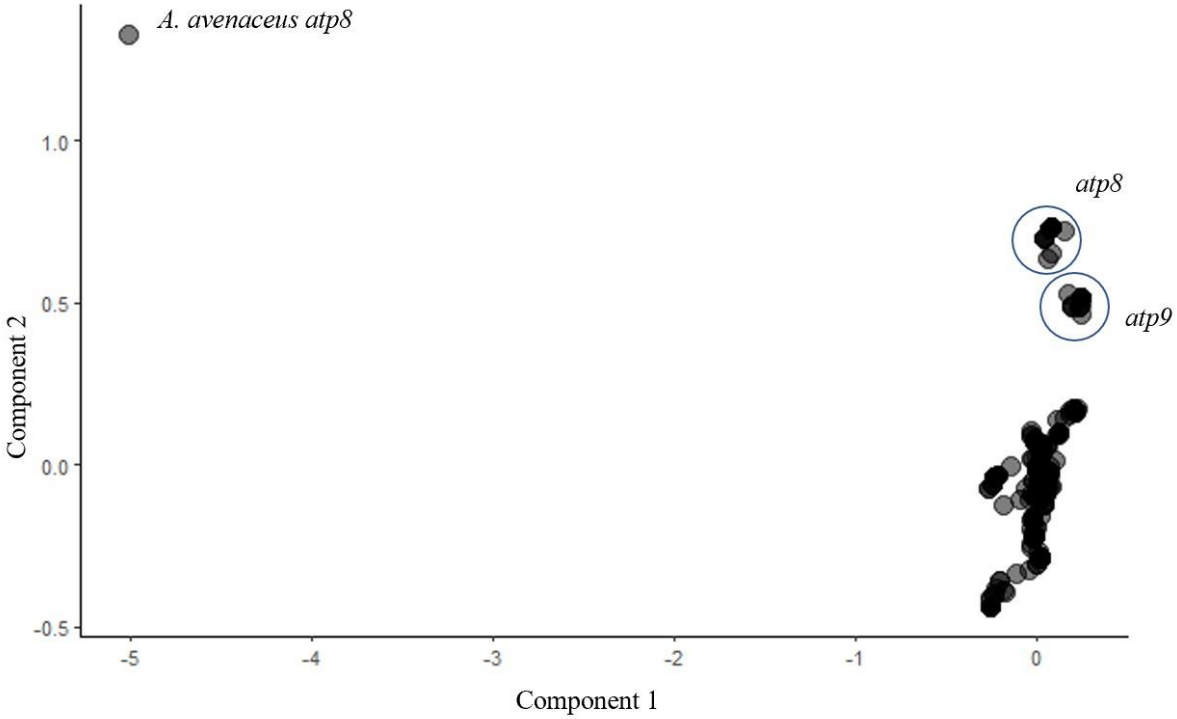
A.



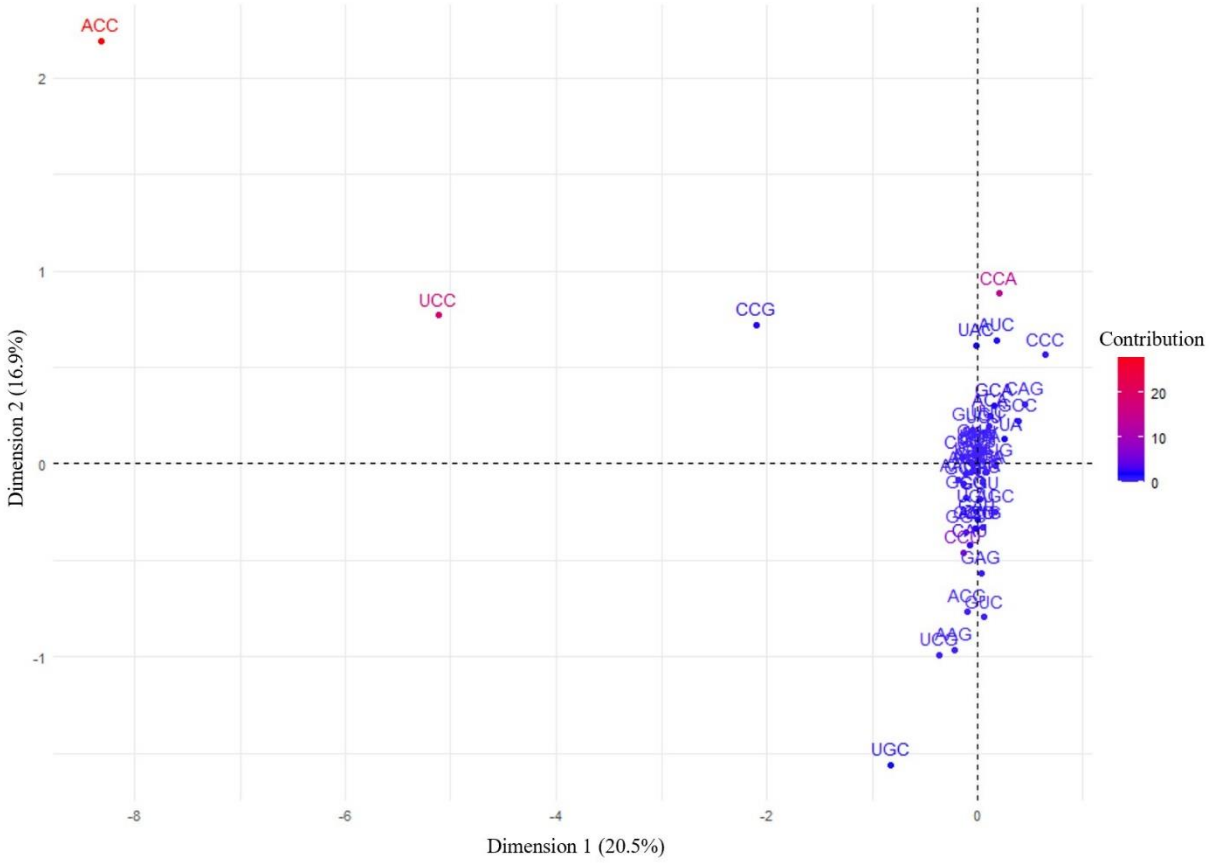
B.

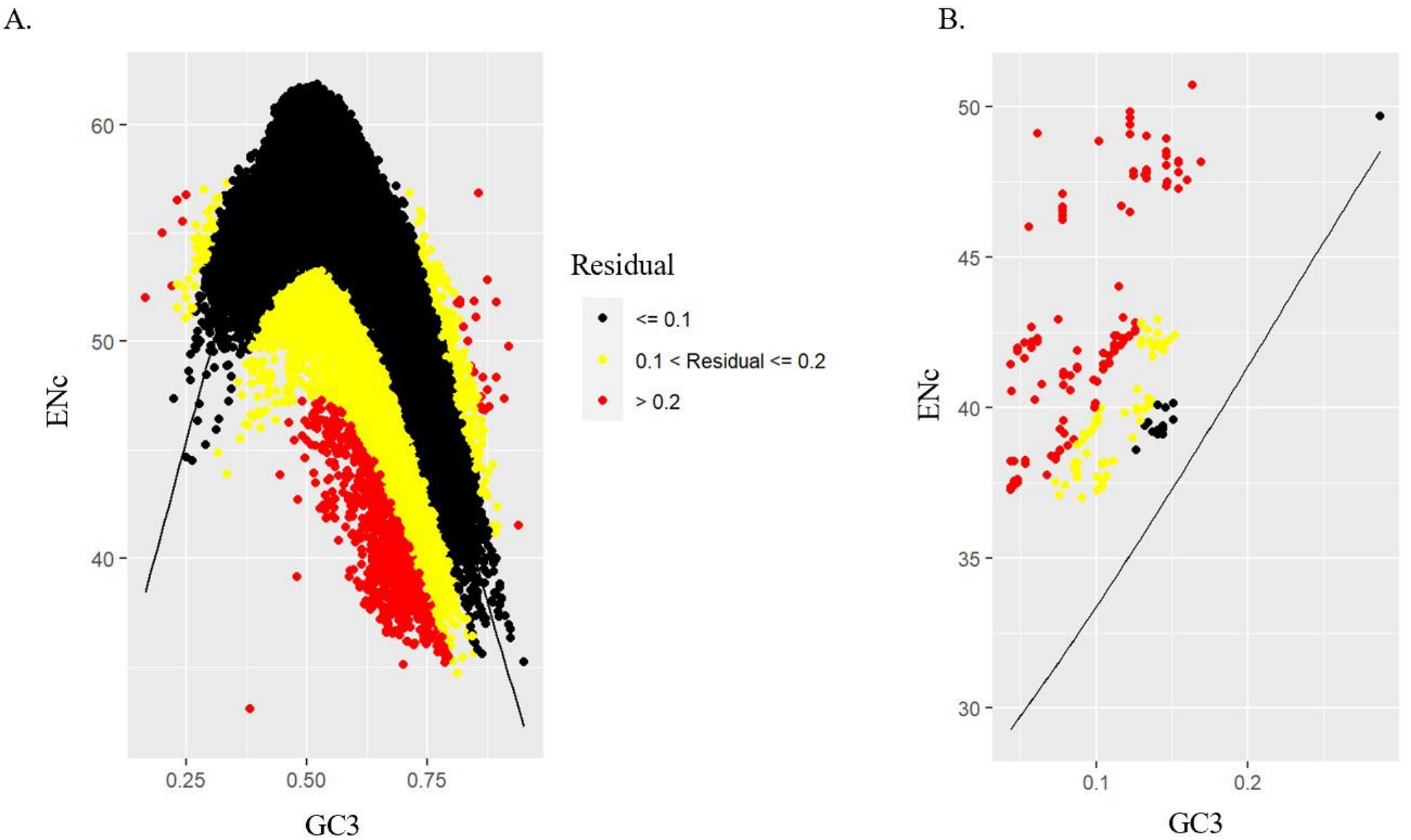


A.

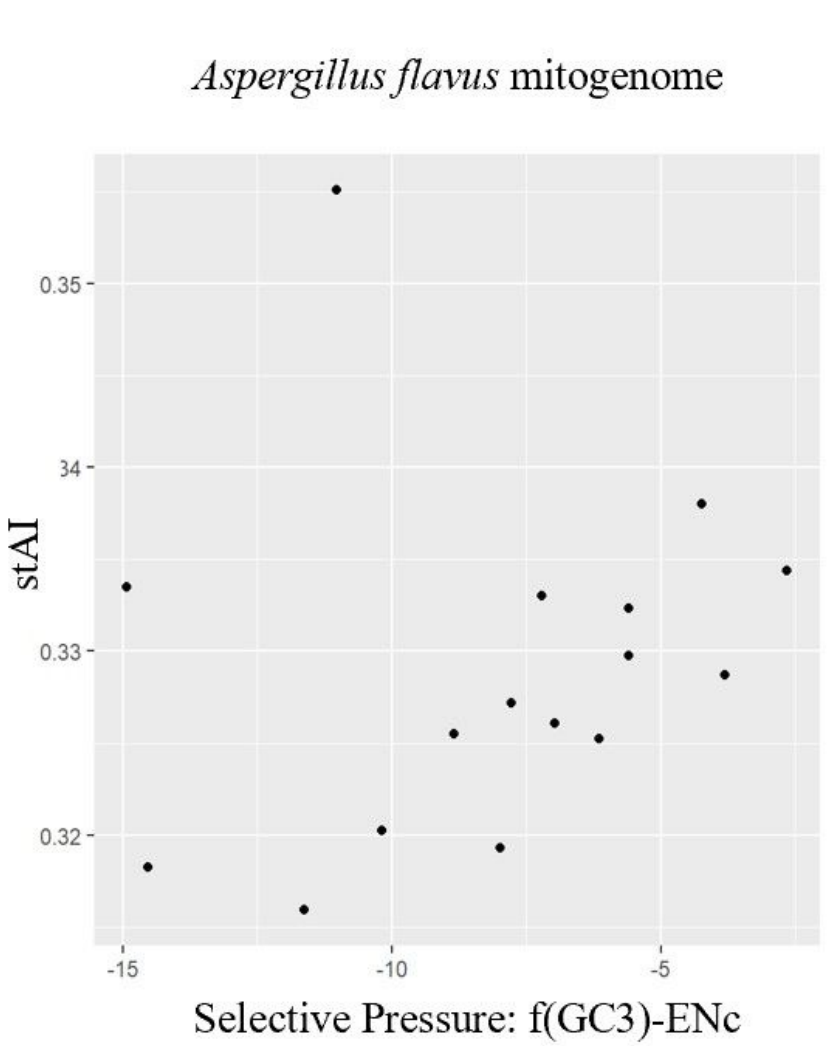


B.





A.



B.