Supporting information

Figure S1:
Relationship between gene family size and production of alternatively spliced variants in human. Fraction of genes containing more than one splice variant is displayed. Binning is similar to Kopelman *et al.* (6) (see also Figure 1A).

Figure S2:
Regression between gene family size and the proportion of genes undergoing alternative splicing in human. Legend similar to Figure 1. A linear regression was tested against a parabola (polynomial regression of order 2; see Methods). Both regression curves are displayed on the graph.

Figure S3:
Relationship between gene family size and the proportion of genes undergoing alternative splicing in mouse (*Mus musculus*). Legend similar to Figure 1.

Figure S4:
Relationship between gene family size and the proportion of genes undergoing alternative splicing in zebrafish (*Danio rerio*). Legend similar to Figure 1.

Figure S5:
Relationship between gene family size and the proportion of genes undergoing alternative splicing in human. Ensembl protein families were used to calculate family size. Legend similar to Figure 1.

Figure S6:
Relationship between gene family size and the proportion of genes undergoing alternative splicing in human. UCSC genome browser was used to estimate number of splice forms per gene. Legend similar to Figure 1.

Figure S7:

Relationship between gene family size and the proportion of genes undergoing alternative splicing in human. Four bins of genes were created using quartiles of EST counts (see Methods) and the trend is displayed for each sub-groups. Legend similar to Figure 1.

Figure S8:
Relationship between gene family size and the proportion of genes undergoing alternative splicing in human. Four bins of genes were created depending on their number of constitutive exons and the trend is displayed for each sub-groups. Genes with no constitutive exon are by definition all undergoing alternative splicing and are not included here. Legend similar to Figure 1.

Figure S9:
Relationship between gene family size and the proportion of genes undergoing alternative splicing in human. Four bins of genes were created using quartiles of d_N/d_S values (calculated with mouse; see Methods) and the trend is displayed for each sub-groups. Legend similar to Figure 1.

Figure S10:
Relationship between gene family size and the proportion of genes undergoing alternative splicing in human. Four bins of genes were created using quartiles of their maximum transcript length and the trend is displayed for each sub-groups. Legend similar to Figure 1.

Figure S11:
Relationship between gene family size and the proportion of genes undergoing alternative splicing in human. Recent duplicates specific to *Homo sapiens* were excluded from the dataset because they might be assembly artefacts. Legend similar to Figure 1.

Figure S12:
Relationship between mean number of alternative spliced variants and age of genes last duplication in human. Legend similar to Figure 2A. The x-axis is in shown linear-scale.

Figure S13:
Relationship between proportion of genes undergoing alternative splicing and age of genes last duplication in huamn. Legend similar to Figure 2.

Figure S14:
Relationship between mean number of alternative spliced variants and age of genes last duplication in human. The analysis was performed separately with the genes mapped to each top-level Gene Ontology functional categories (see Methods). Legend similar to Figure 2.

Figure S15:
Relationship between the difference of number of alternative spliced variants for gene duplicate pairs and age of genes last duplication in human. The analysis was restricted to duplicate pairs for which no later duplication occurred (see Methods). Legend similar to Figure 2.

Figure S16:
Histograms comparing the number of splice forms between genes whose last duplication is old (Bilateria; in black) and genes whose last duplication is recent (*Homo sapiens*; in gray).

Figure S17:
Relationship between proportion of singleton genes undergoing alternative splicing and their age in human. Legend similar to Figure 2.

Figure S18:
Relationship between mean number of alternative spliced variants and age of singletons in human. Legend similar to Figure 2B. The x-axis is in shown linear-scale.

Figure S19:
Relationship between mean number of alternative spliced variants and age of singletons in human. Four bins of genes were created using quartiles of d_N/d_S values (calculated with mouse; see Methods) and the trend is displayed for each sub-groups. Legend similar to Figure 2.

Figure S20:
(A) Comparison of proportion of genes undergoing alternative splicing for groups of genes with different evolutionary histories in human. “Duplicates” are genes that kept in duplicate after the whole-genome duplications ancestral to vertebrates (2R; later duplications allowed).

“Singletons” are genes whose duplicate was not retained after 2R whole-genome duplications (later duplications not allowed). “Singletons (+ later duplications)” represent genes whose duplicate was not retained after 2R whole-genome duplications, but later duplications were experienced.
(B) Comparison of mean number of alternative spliced variants for groups of genes with different evolutionary histories in human.
Legend similar to Figure 1.

Figure S21:
Relationship between mean number of alternative spliced variants and age of genes last duplication in human. Only Ensembl transcripts encoding for a protein were kept in the dataset. Legend similar to Figure 2.

Figure S22:
(A) Relationship between proportion of genes undergoing alternative splicing and age of genes last duplication in human. Only alternative splice forms annotated by Swiss-Prot as “experimentally validated” were kept in the dataset.
(B) Relationship between mean number of alternative spliced variants and age of genes last duplication in human. Only alternative splice forms annotated by Swiss-Prot as “experimentally validated” were kept in the dataset. Legend similar to Figure 2.

Figure S23:
(A) Relationship between proportion of genes undergoing alternative splicing and age of genes last duplication in human. Only alternative splice forms conserved with mouse (extracted from the H-DBAS database) were kept in the dataset.
(B) Relationship between mean number of alternative spliced variants and age of genes last duplication in human. Only alternative splice forms conserved with mouse (extracted from the H-DBAS database) were kept in the dataset. Legend similar to Figure 2.

Figure S24:
Comparison of linear regression models explaining the mean number of alternative spliced variants by the age of singletons in human. Data were split in sub-groups and the curves are displayed for each sub-group (see legend on the graph).

(A) Linear models were adjusted on four bins of genes depending on number of constitutive exons (see Figure S8).
(B) Linear models were adjusted on four bins of genes depending on maximum transcript length (see Figure S10).
(C) Linear models were adjusted on four bins of genes depending on d_N/d_S values (see Figures S9 and S19).
(D) Linear models were adjusted on four bins of genes depending on EST counts (see Figure S7).
(E) Linear models were adjusted on four datasets where only splice forms corresponding to a given type of splicing events (e.g., cassette exons, CE) were kept.

Figure S25:
(A) Relationship between mean number of alternative spliced variants and the age of human genes whose last duplication occurred in the branch leading to *Homo sapiens*.
(B) Relationship between mean number of alternative spliced variants and the age of human genes whose last duplication occurred in the branch leading to Eutheria.
Legend similar to Figure 2.

Figure S26:
Relationship between mean rate of sequence evolution (mean log d_N/d_S) of singleton genes and their age in human. Legend similar to Figure 2.

Figure S27:
Relationship between gene age and production of alternatively spliced variants in *Drosophila melanogaster*.
(A) Legend as in Figure 2A.
(B) Legend as in Figure S17.

Dataset S1:
Excel file gathering the dataset used for this study.

Figure S1

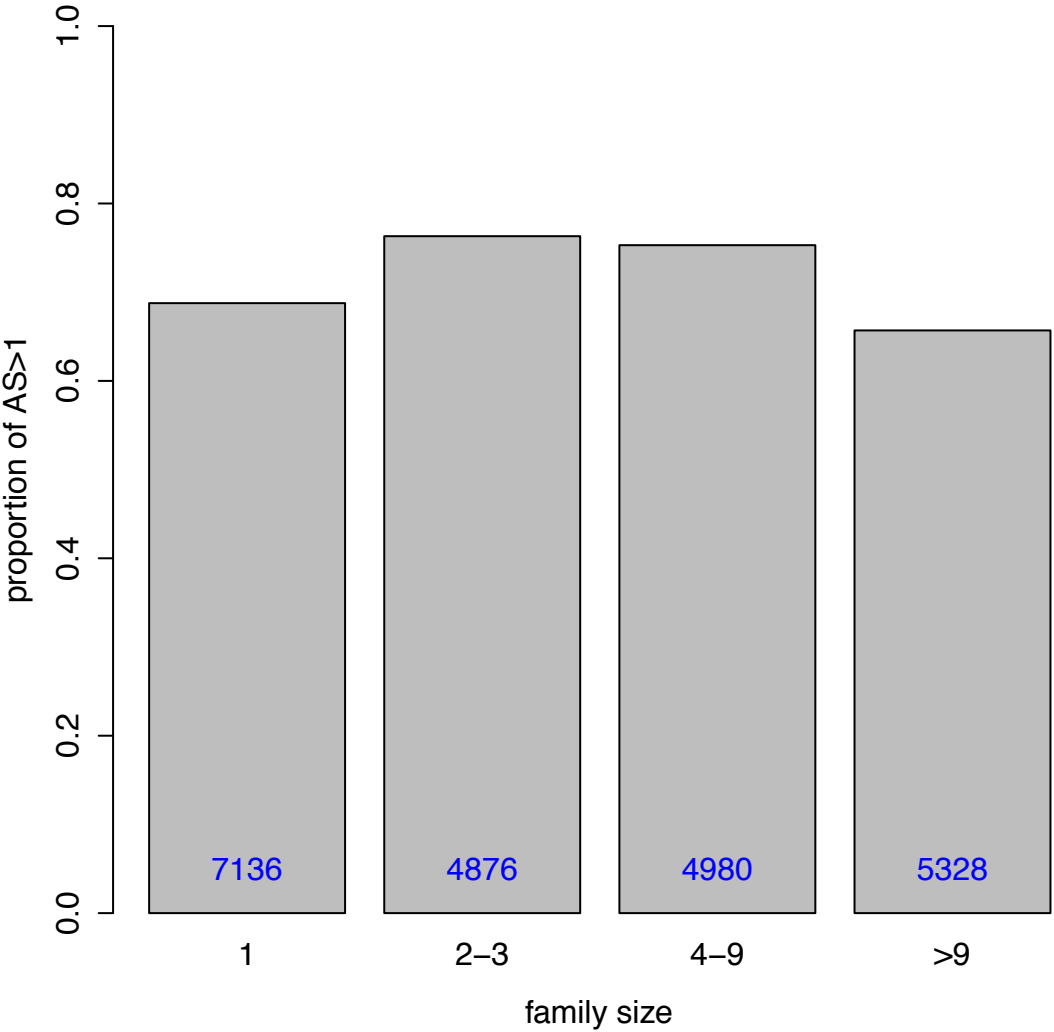

proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
7136
4876
4980
5328
1
2–3
4–9
>9
family size

Figure S2

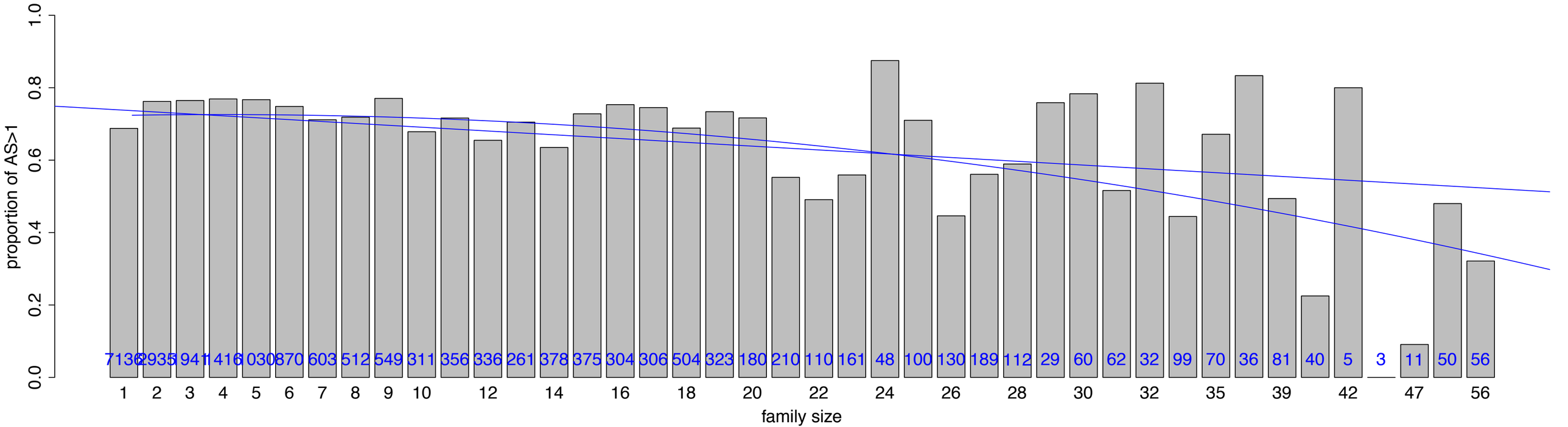

proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
7136
2935
941
416
030
870
603
512
549
311
356
336
261
378
375
304
306
504
323
180
210
110
161
48
100
130
189
112
29
60
62
32
99
70
36
81
40
5
3
11
50
56
1
2
3
4
5
6
7
8
9
10
12
14
16
18
20
22
24
26
28
30
32
35
39
42
47
56
family size

Figure S3

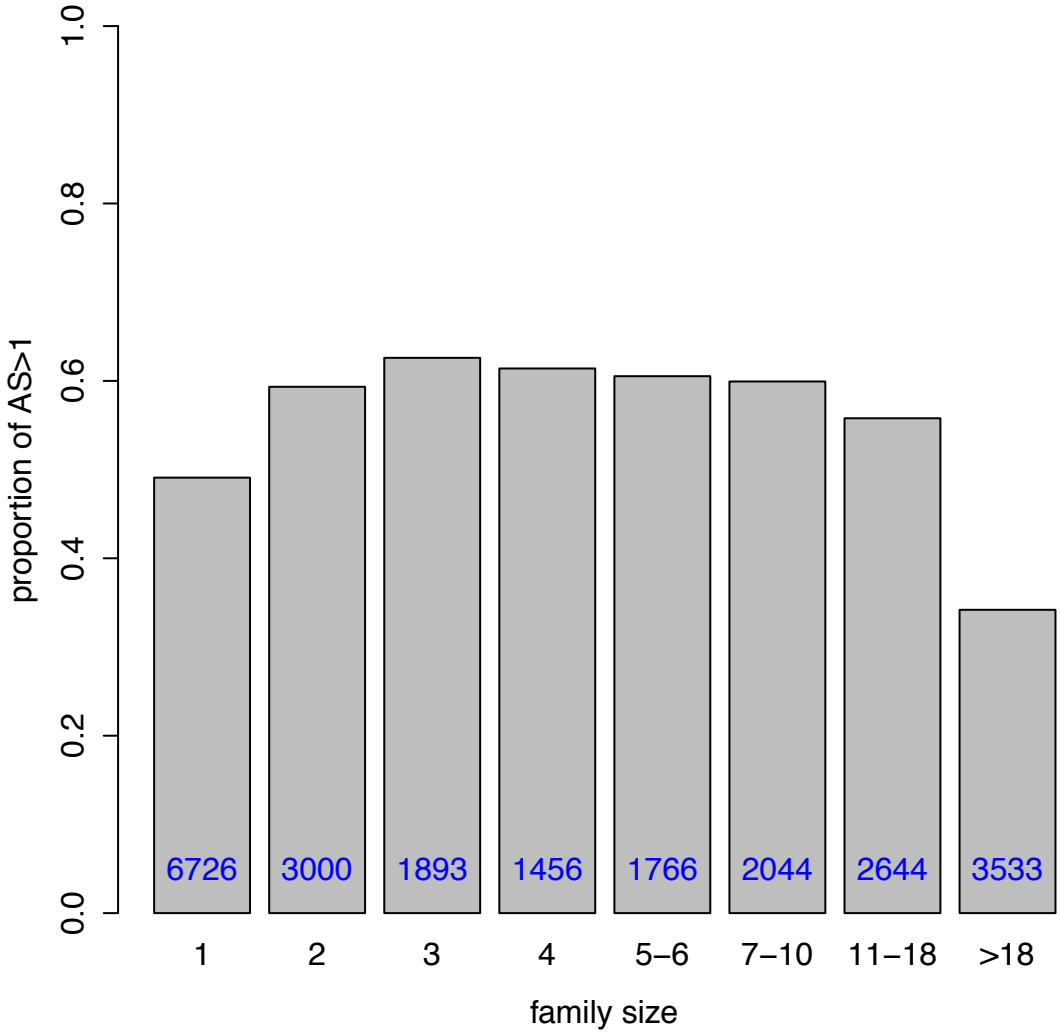

proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
6726
3000
1893
1456
1766
2044
2644
3533
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S4

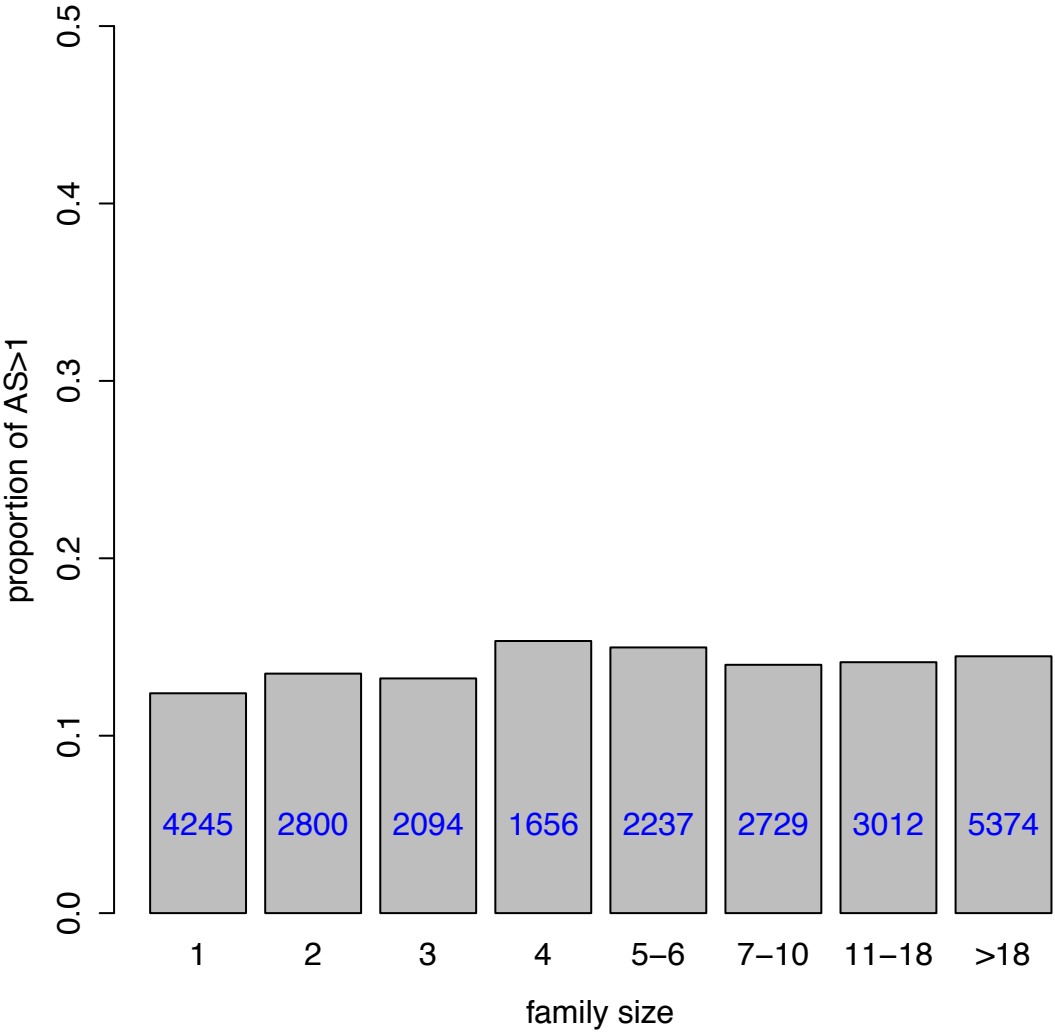

proportion of AS>1
0.0
0.1
0.2
0.3
0.4
0.5
4245
2800
2094
1656
2237
2729
3012
5374
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S5

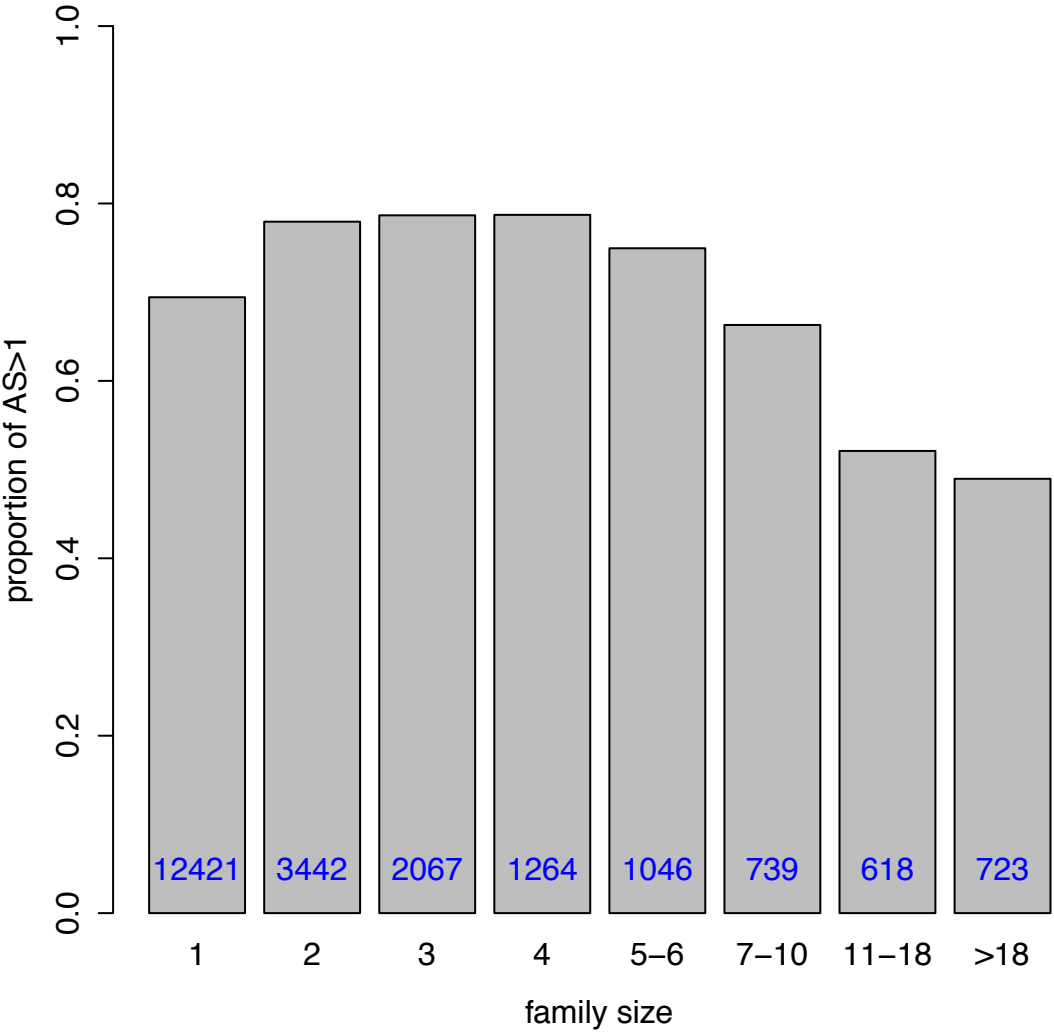

proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
12421
3442
2067
1264
1046
739
618
723
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S6

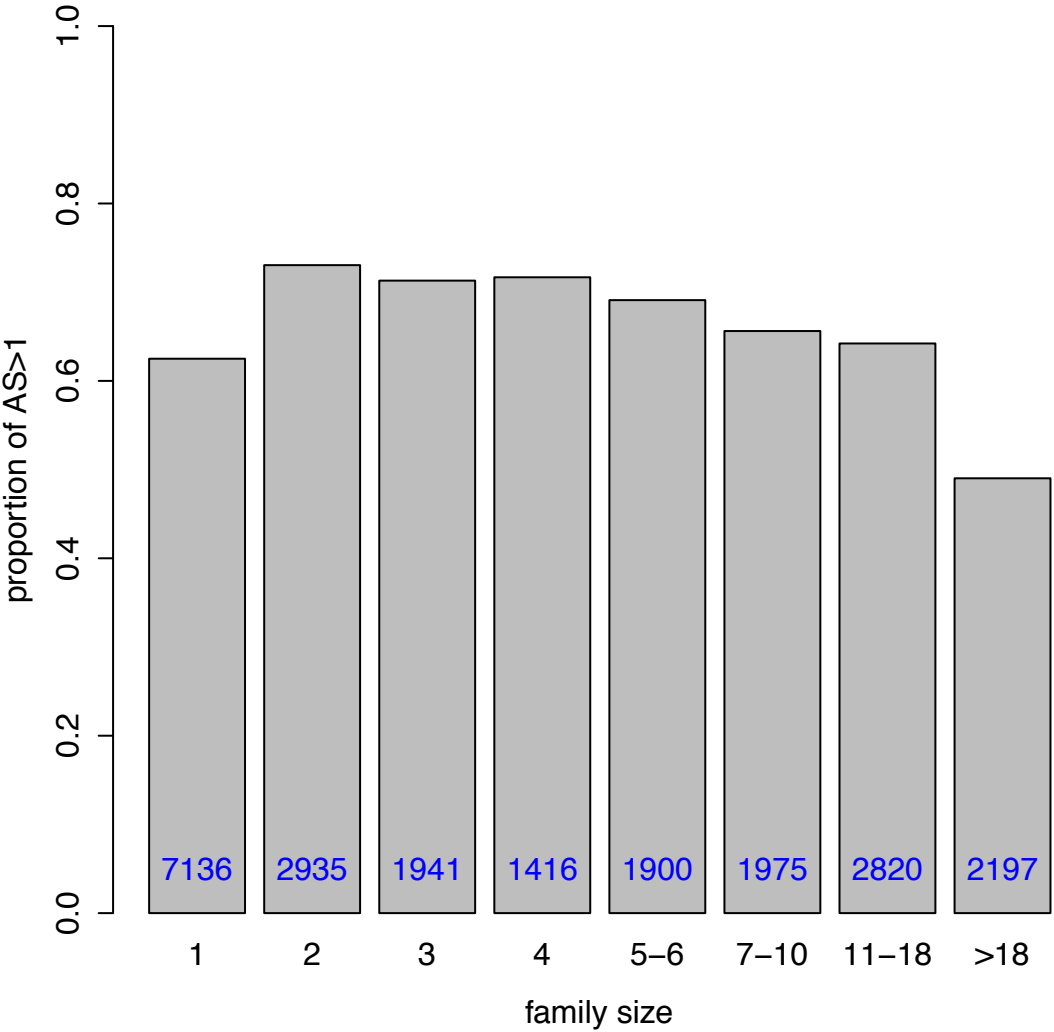
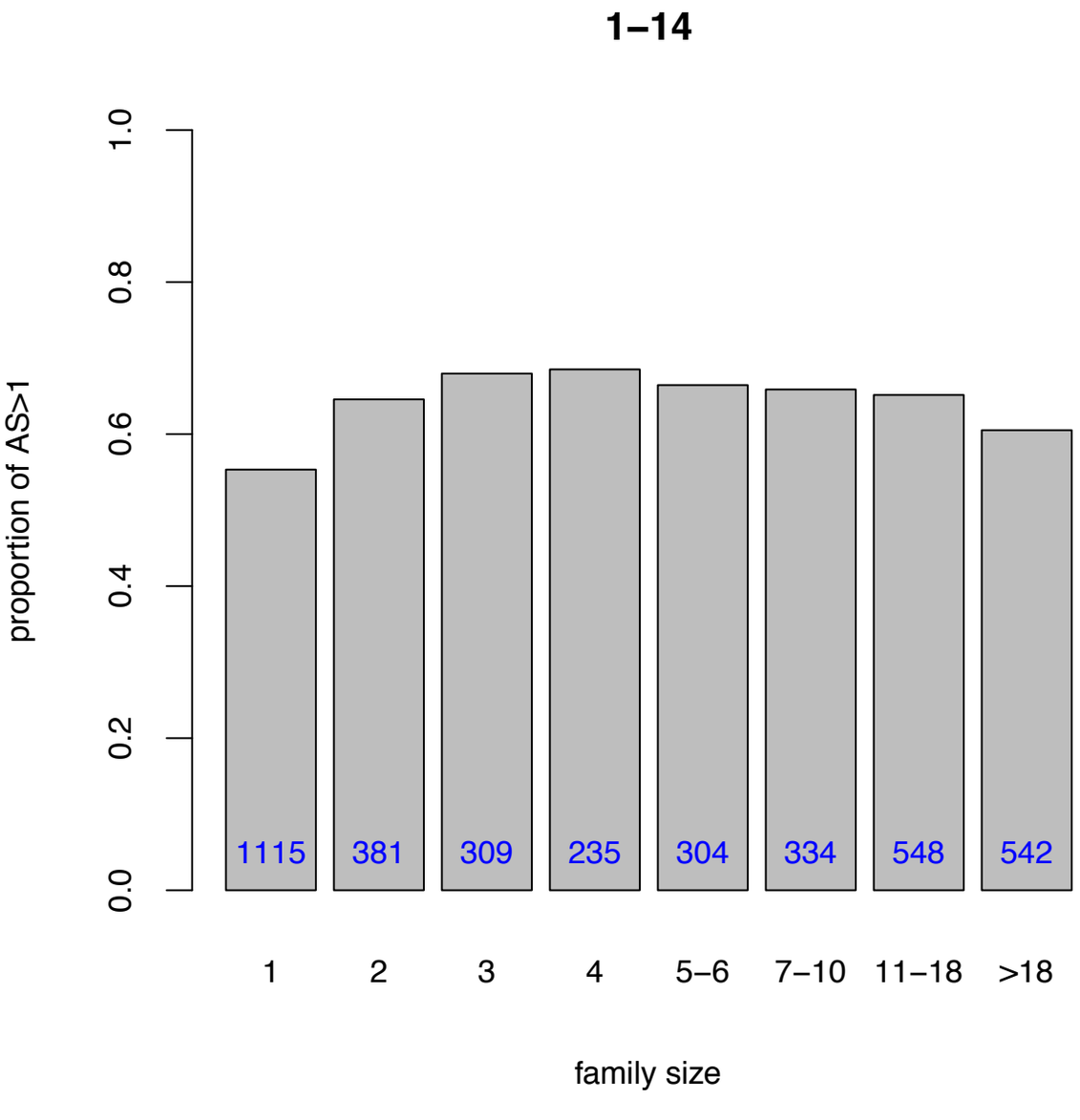
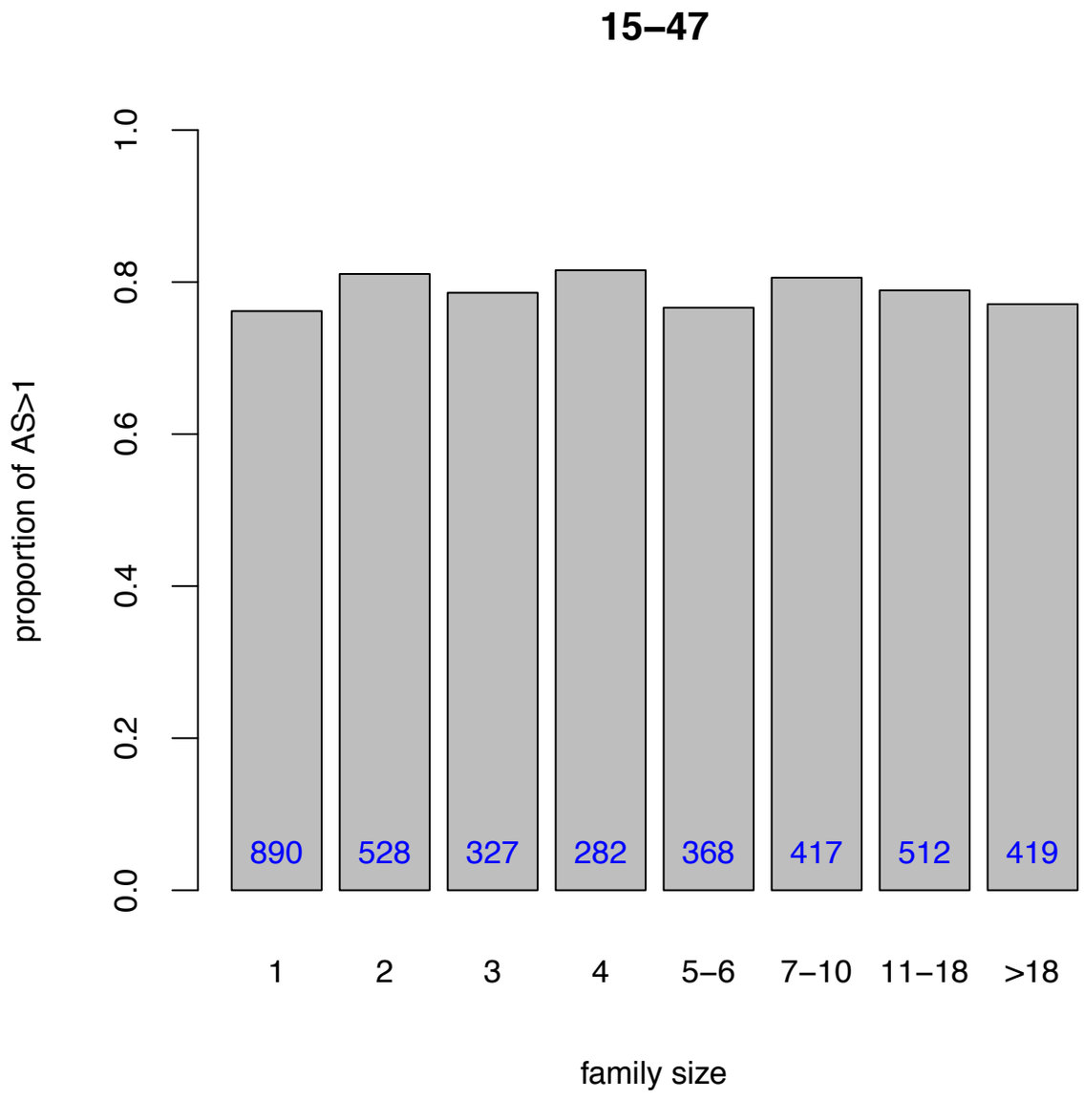
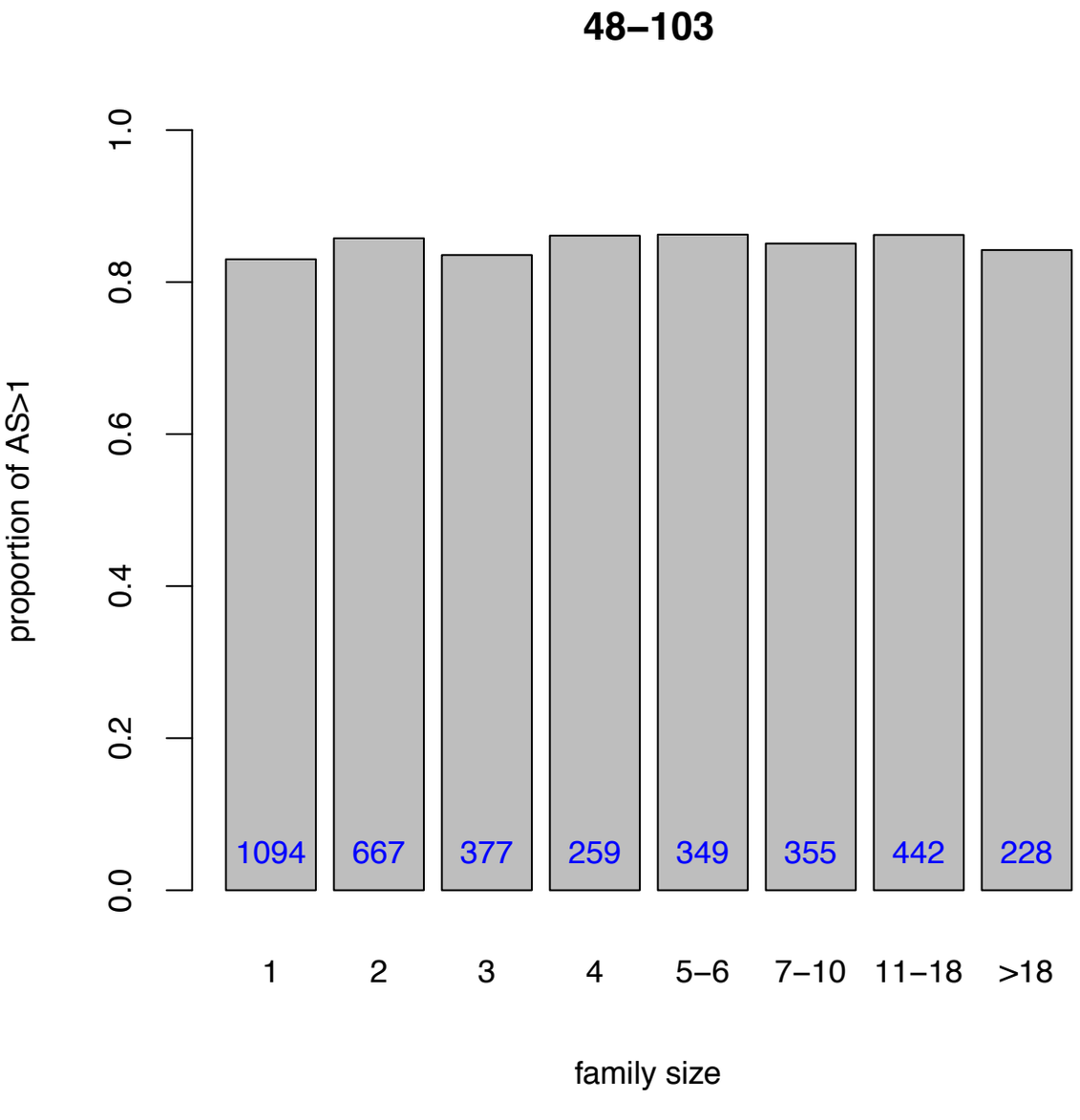

proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
7136
2935
1941
1416
1900
1975
2820
2197
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S7


1–14
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
1115
381
309
235
304
334
548
542
1
2
3
4
5–6
7–10
11–18
>18
family size


15–47
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
890
528
327
282
368
417
512
419
1
2
3
4
5–6
7–10
11–18
>18
family size


48–103
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
1094
667
377
259
349
355
442
228
1
2
3
4
5–6
7–10
11–18
>18
family size

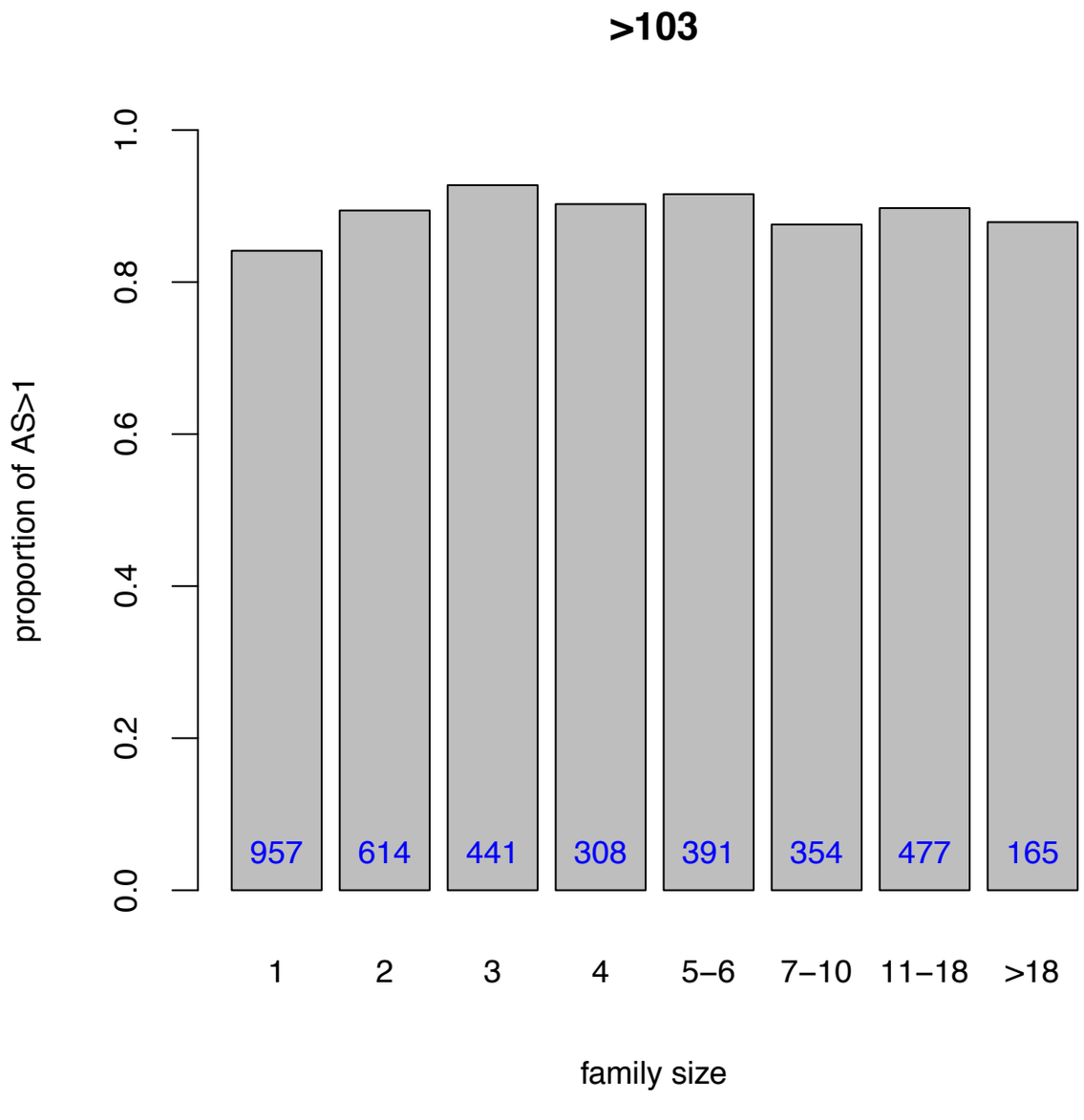
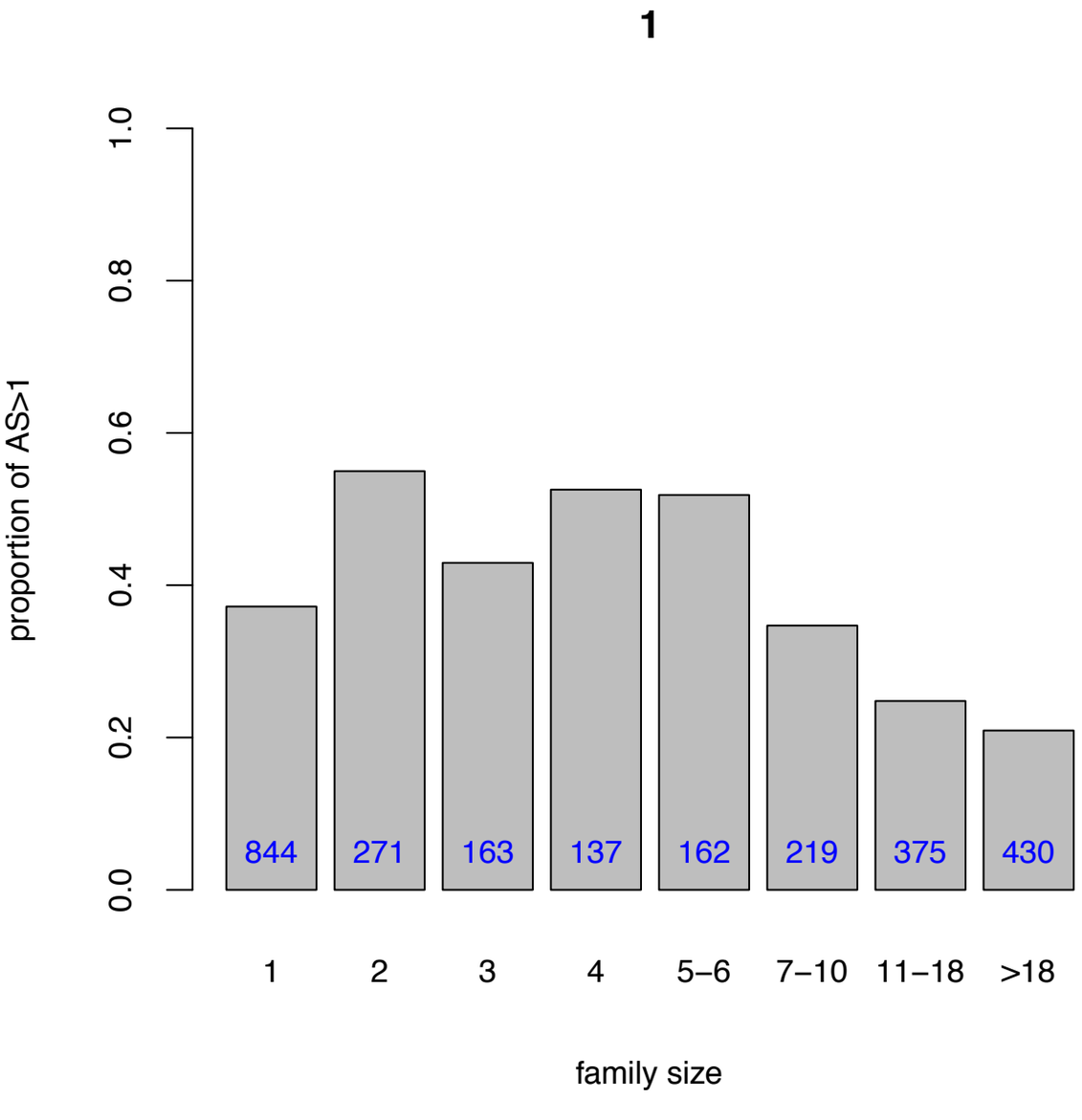
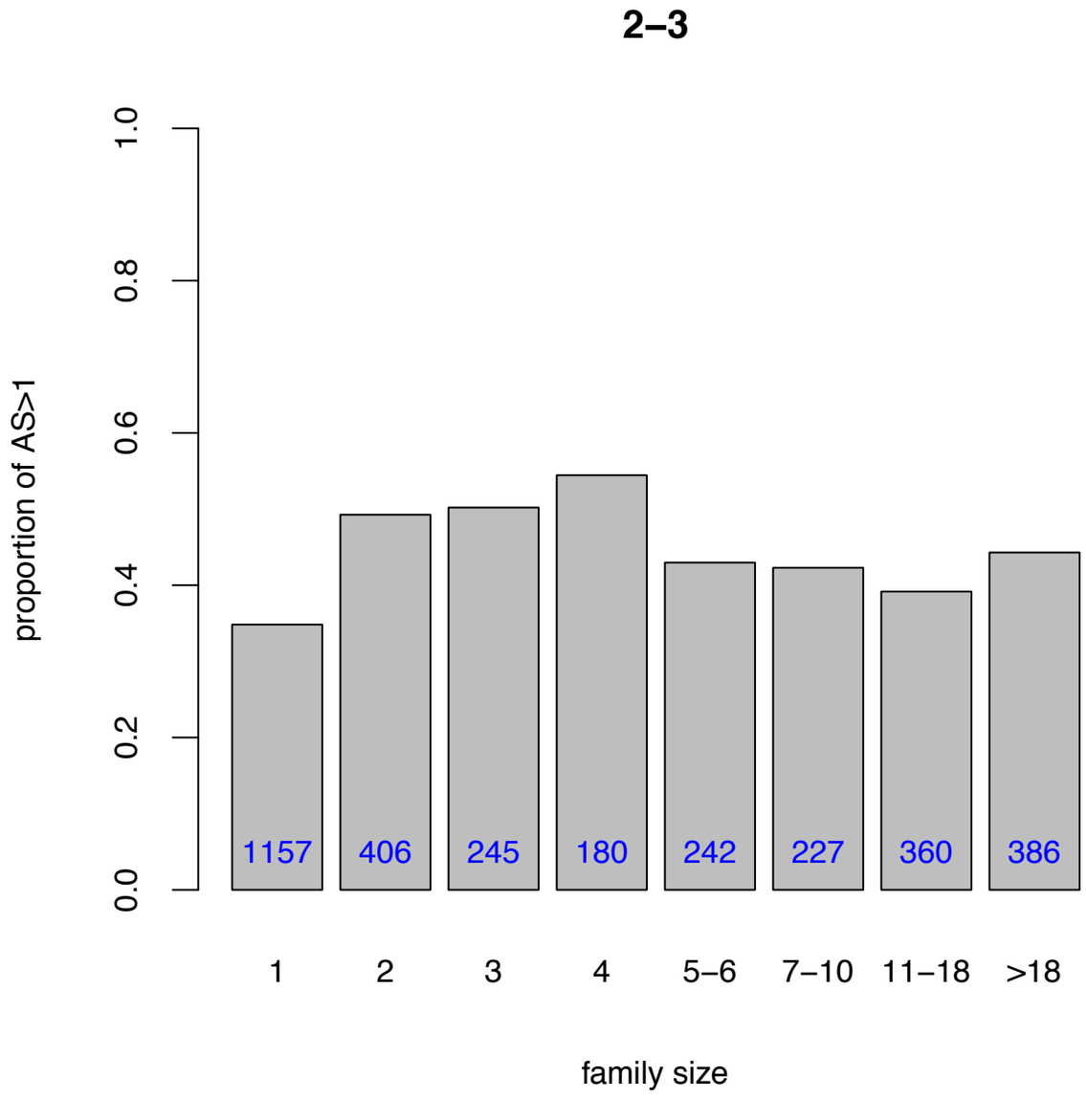
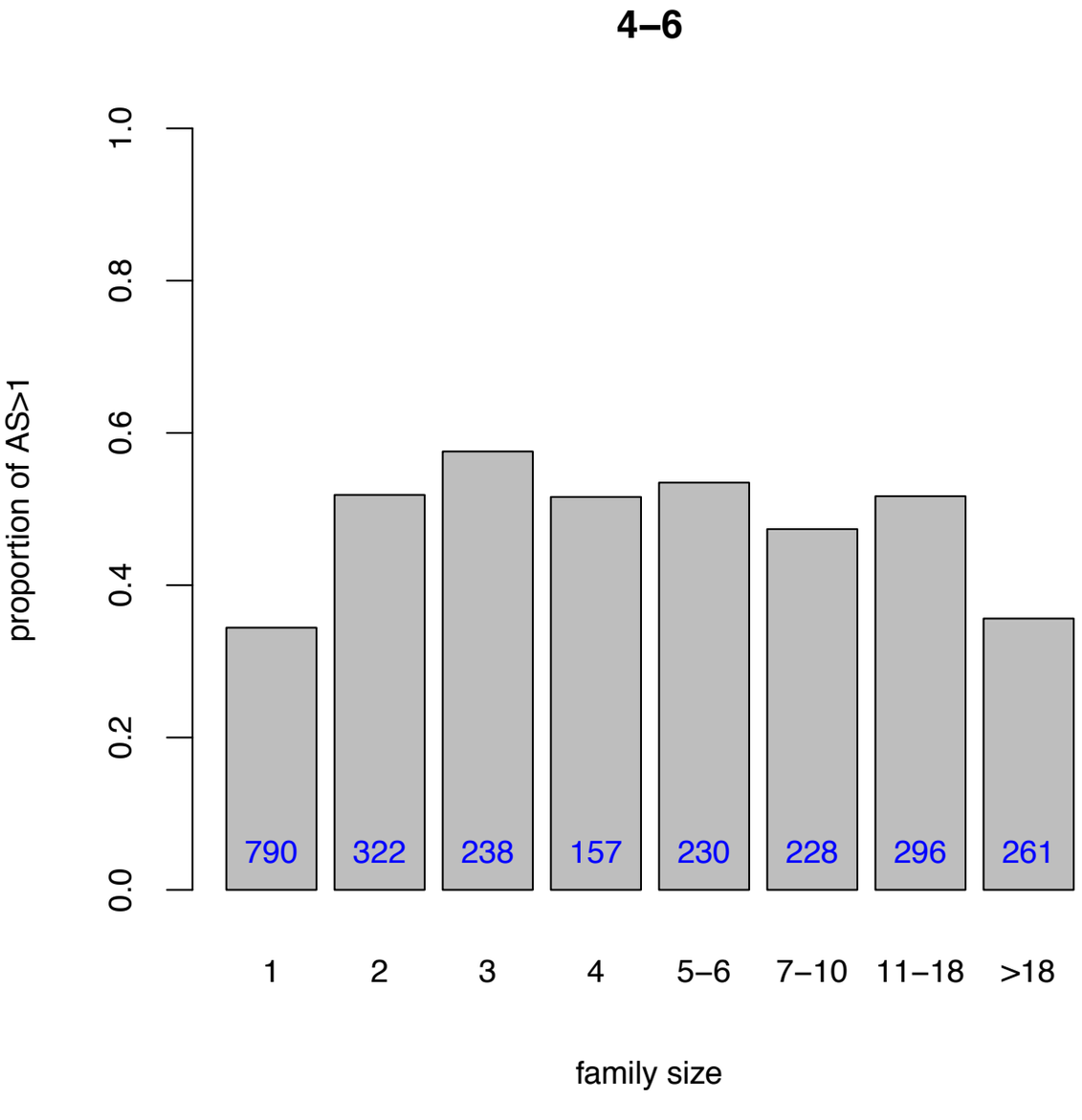

>103
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
957
614
441
308
391
354
477
165
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S8


1
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
844
271
163
137
162
219
375
430
1
2
3
4
5–6
7–10
11–18
>18
family size


2–3
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
1157
406
245
180
242
227
360
386
1
2
3
4
5–6
7–10
11–18
>18
family size


4–6
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
790
322
238
157
230
228
296
261
1
2
3
4
5–6
7–10
11–18
>18
family size

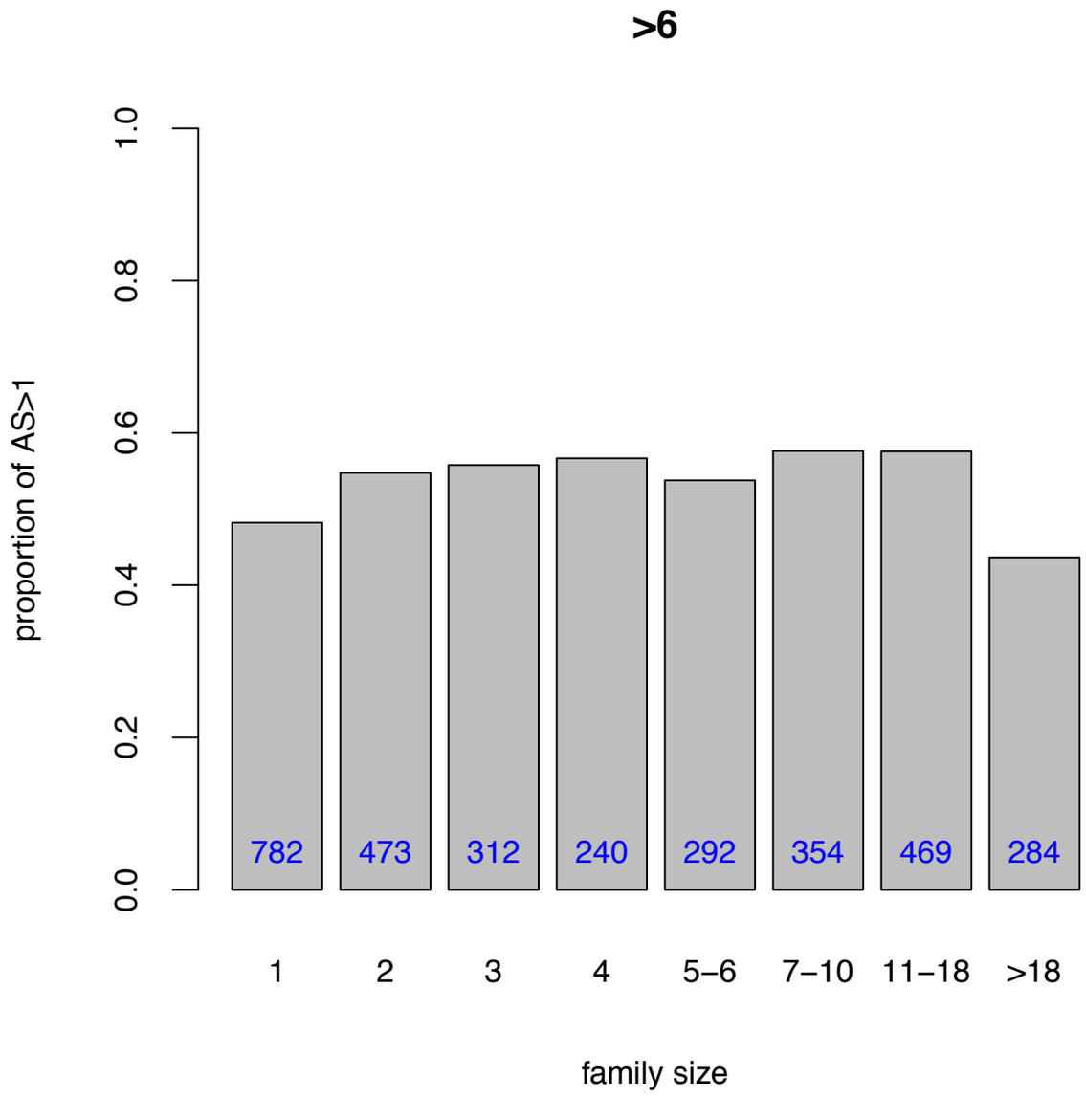
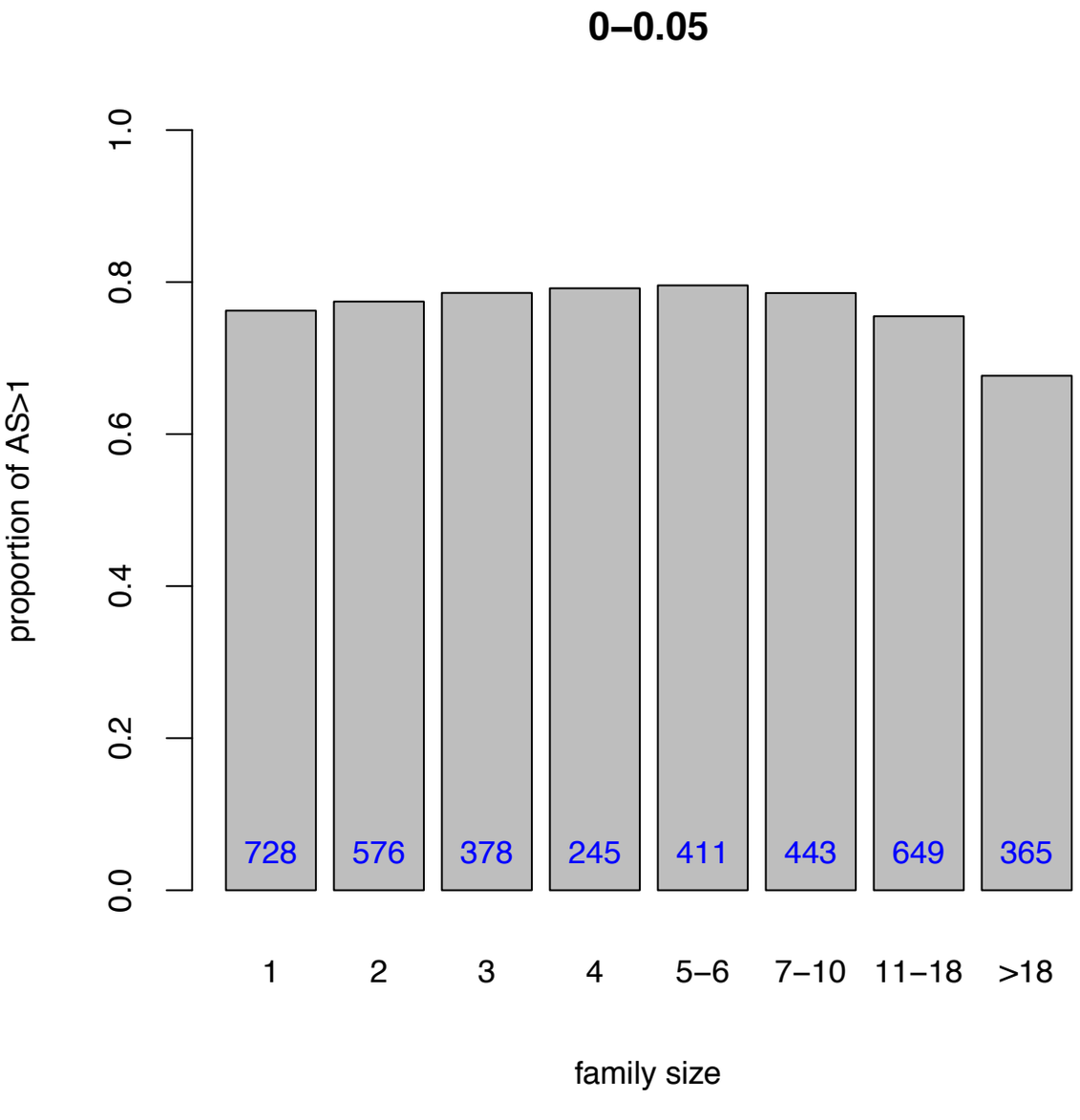
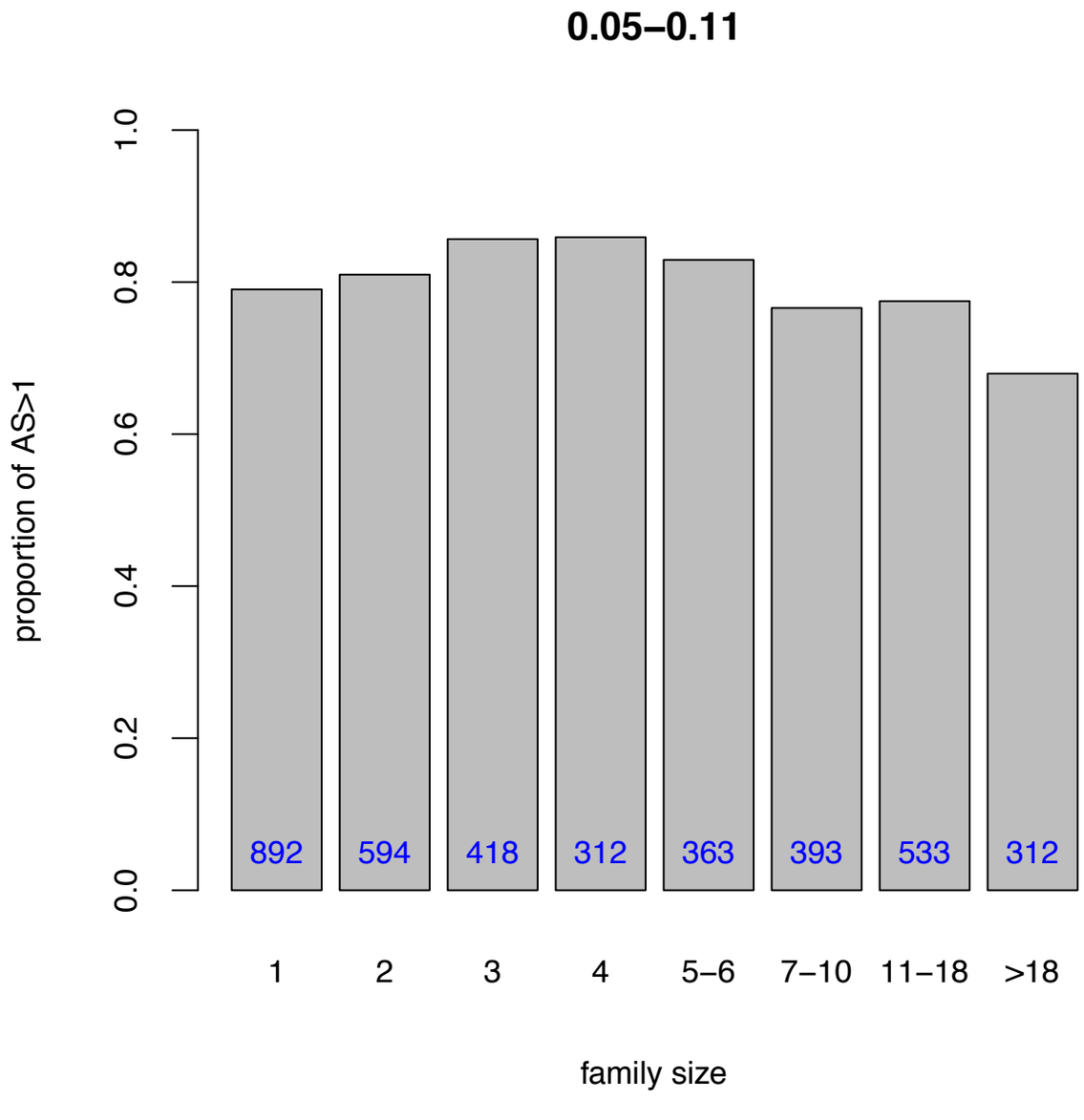
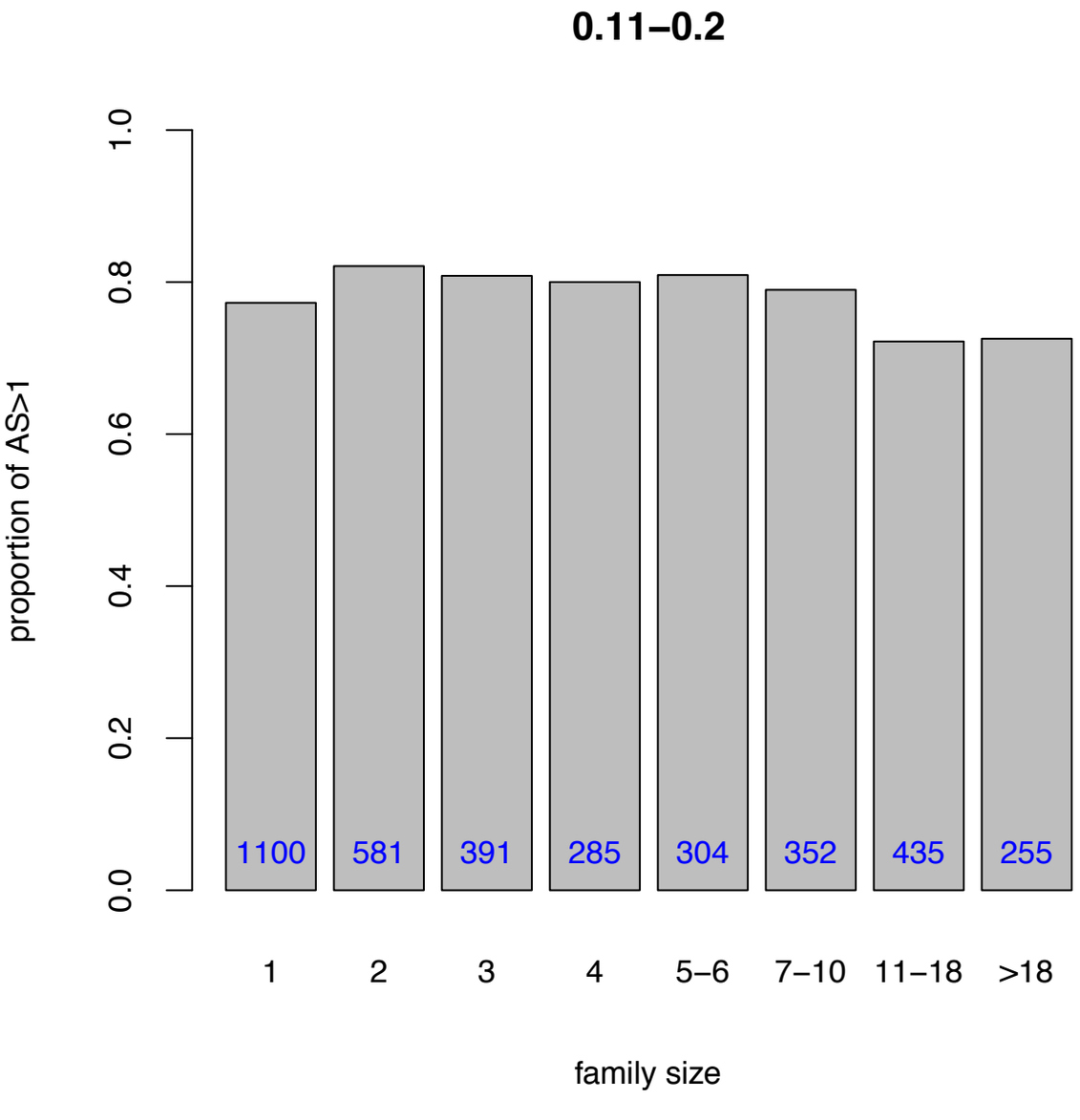

>6
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
782
473
312
240
292
354
469
284
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S9


0–0.05
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
728
576
378
245
411
443
649
365
1
2
3
4
5–6
7–10
11–18
>18
family size


0.05–0.11
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
892
594
418
312
363
393
533
312
1
2
3
4
5–6
7–10
11–18
>18
family size


0.11–0.2
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
1100
581
391
285
304
352
435
255
1
2
3
4
5–6
7–10
11–18
>18
family size

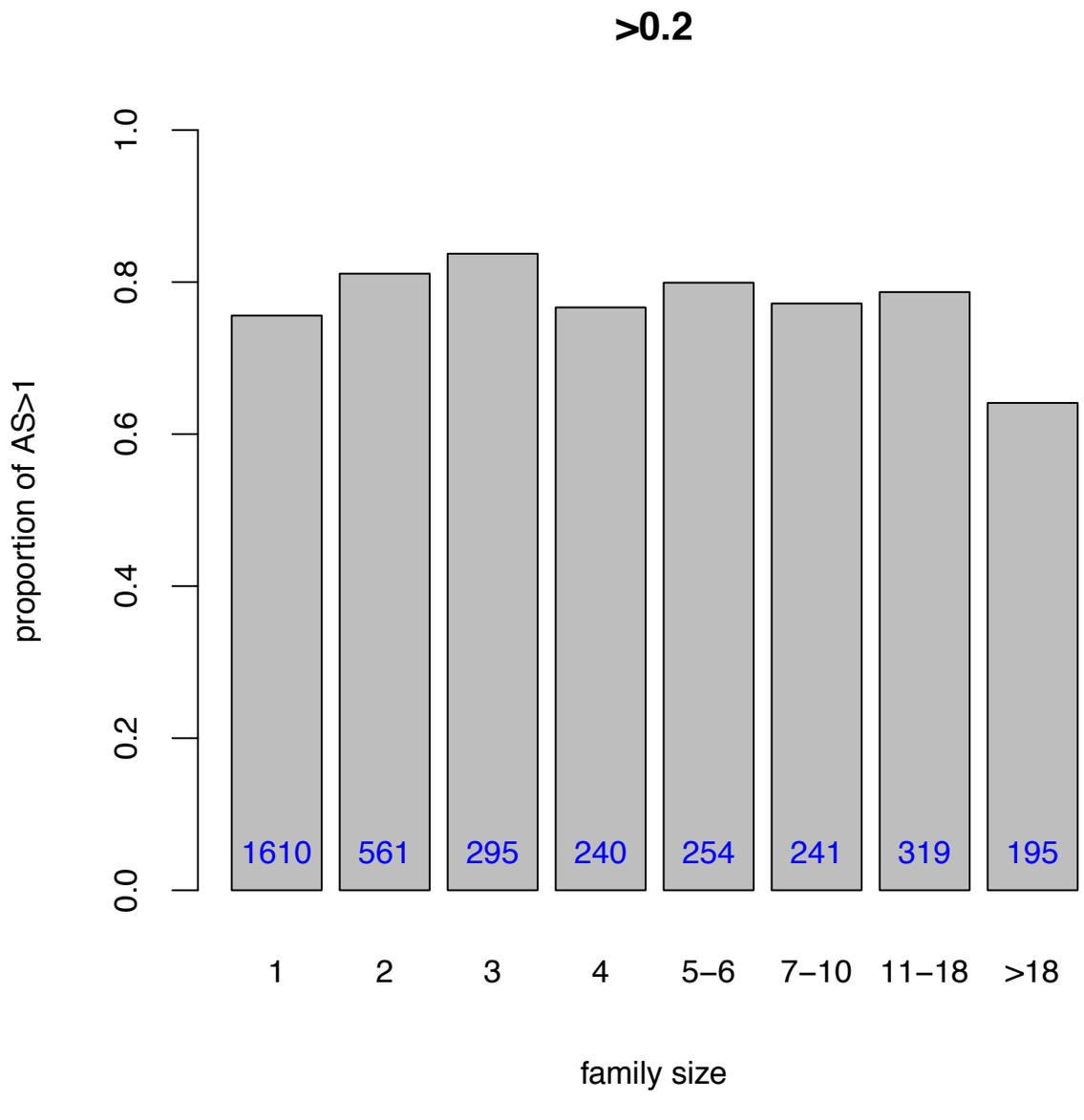
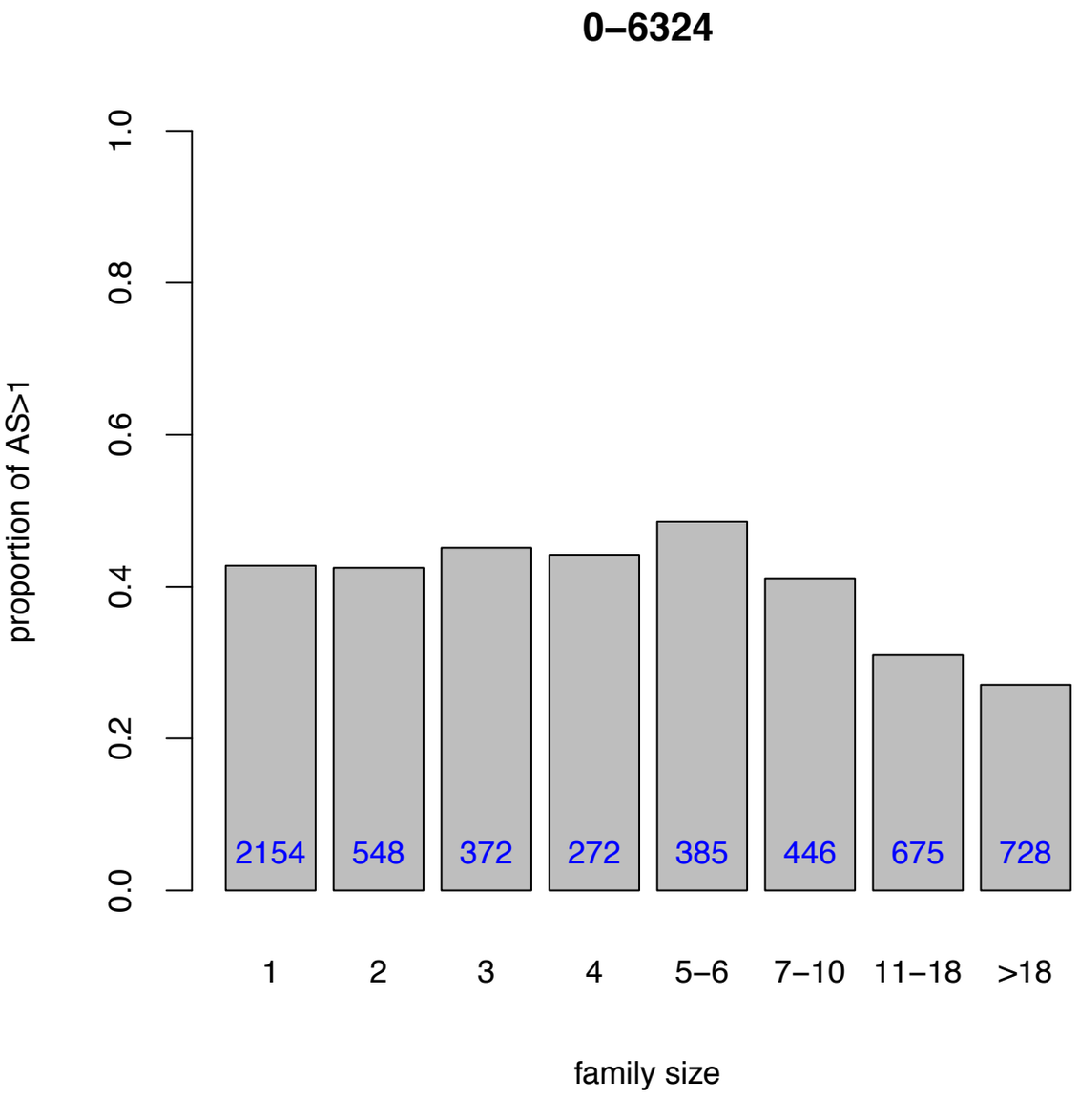
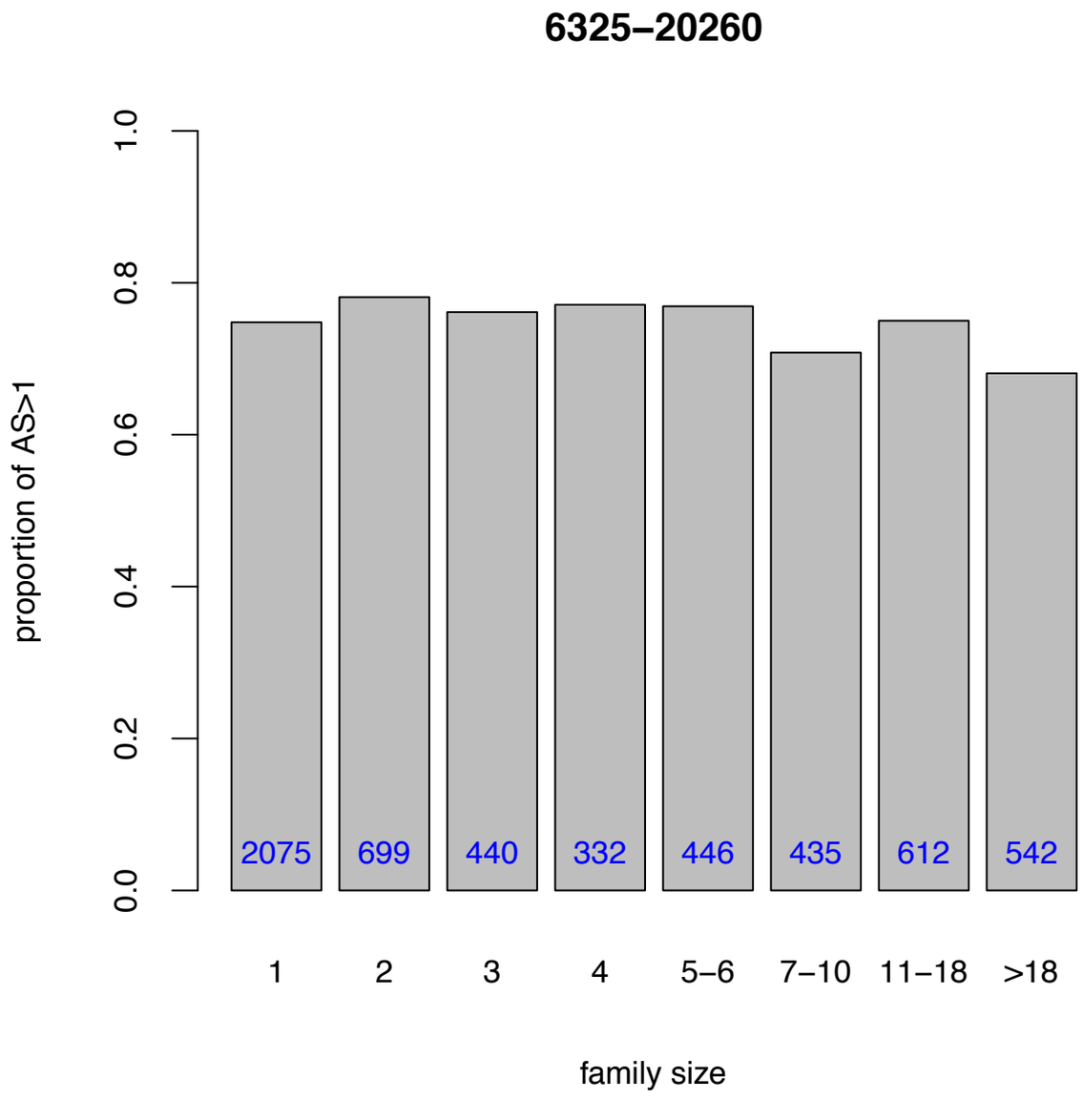
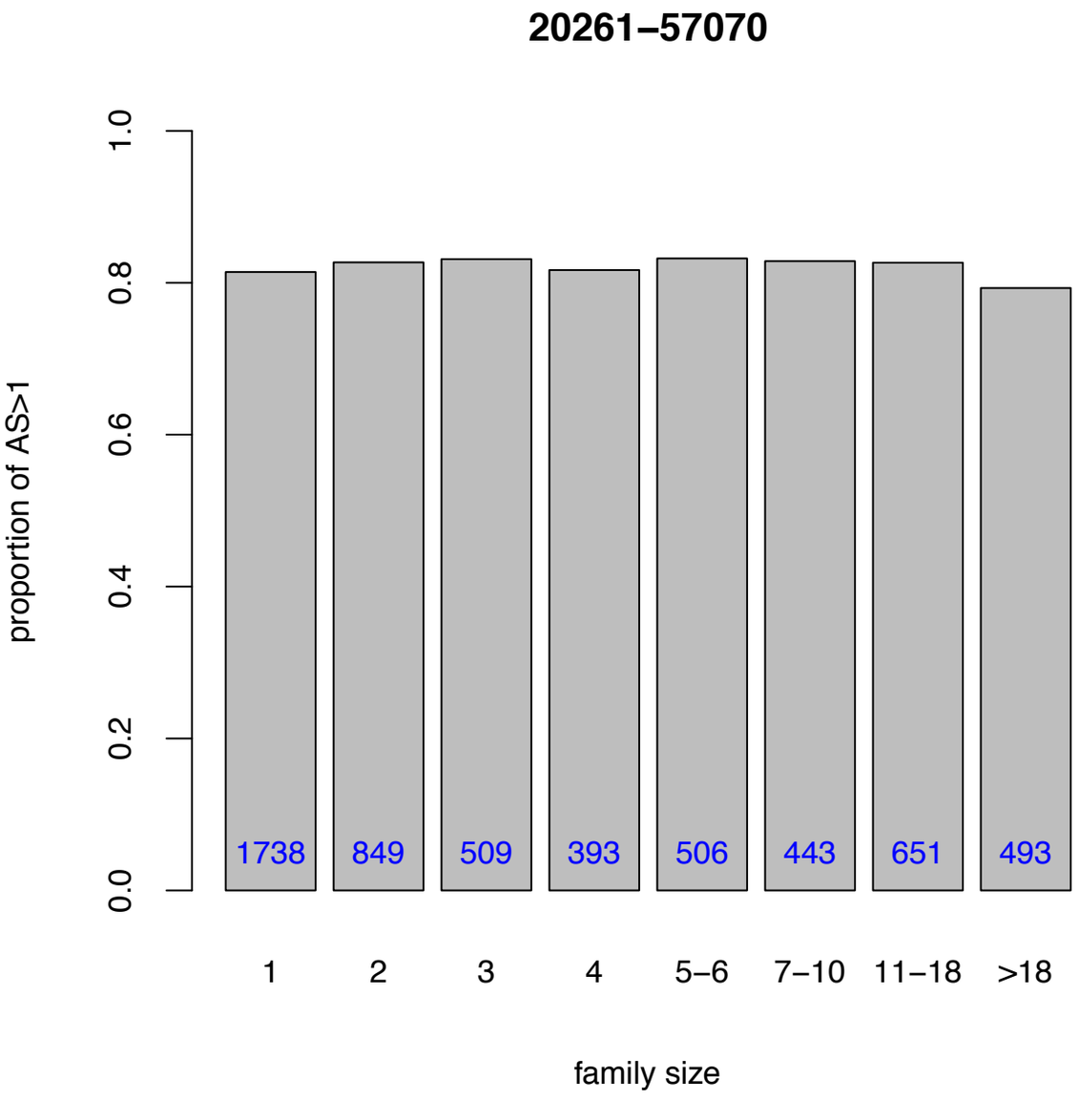

>0.2
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
1610
561
295
240
254
241
319
195
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S10


0–6324
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
2154
548
372
272
385
446
675
728
1
2
3
4
5–6
7–10
11–18
>18
family size


6325–20260
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
2075
699
440
332
446
435
612
542
1
2
3
4
5–6
7–10
11–18
>18
family size


20261–57070
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
1738
849
509
393
506
443
651
493
1
2
3
4
5–6
7–10
11–18
>18
family size

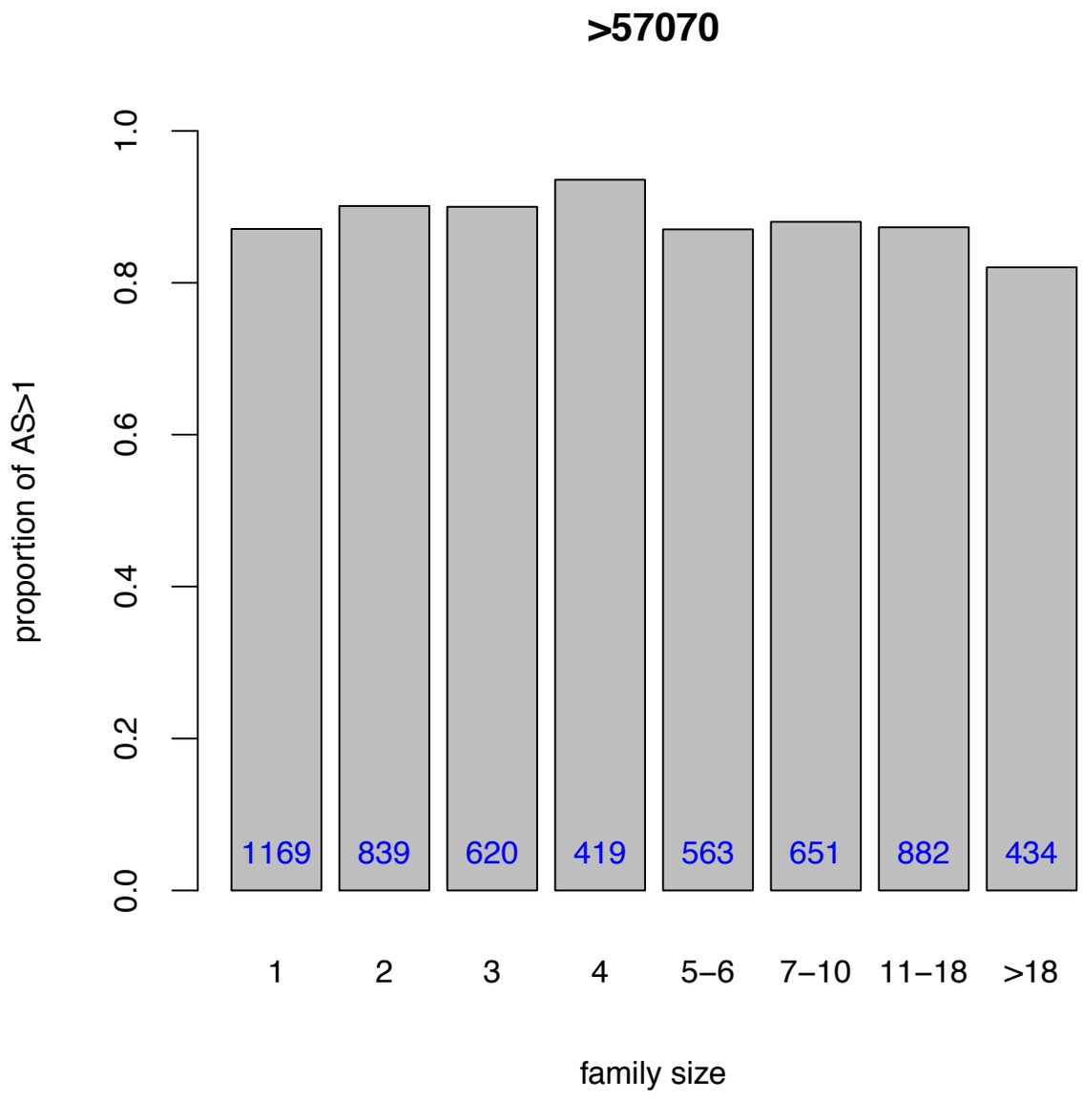

>57070
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
1169
839
620
419
563
651
882
434
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S11

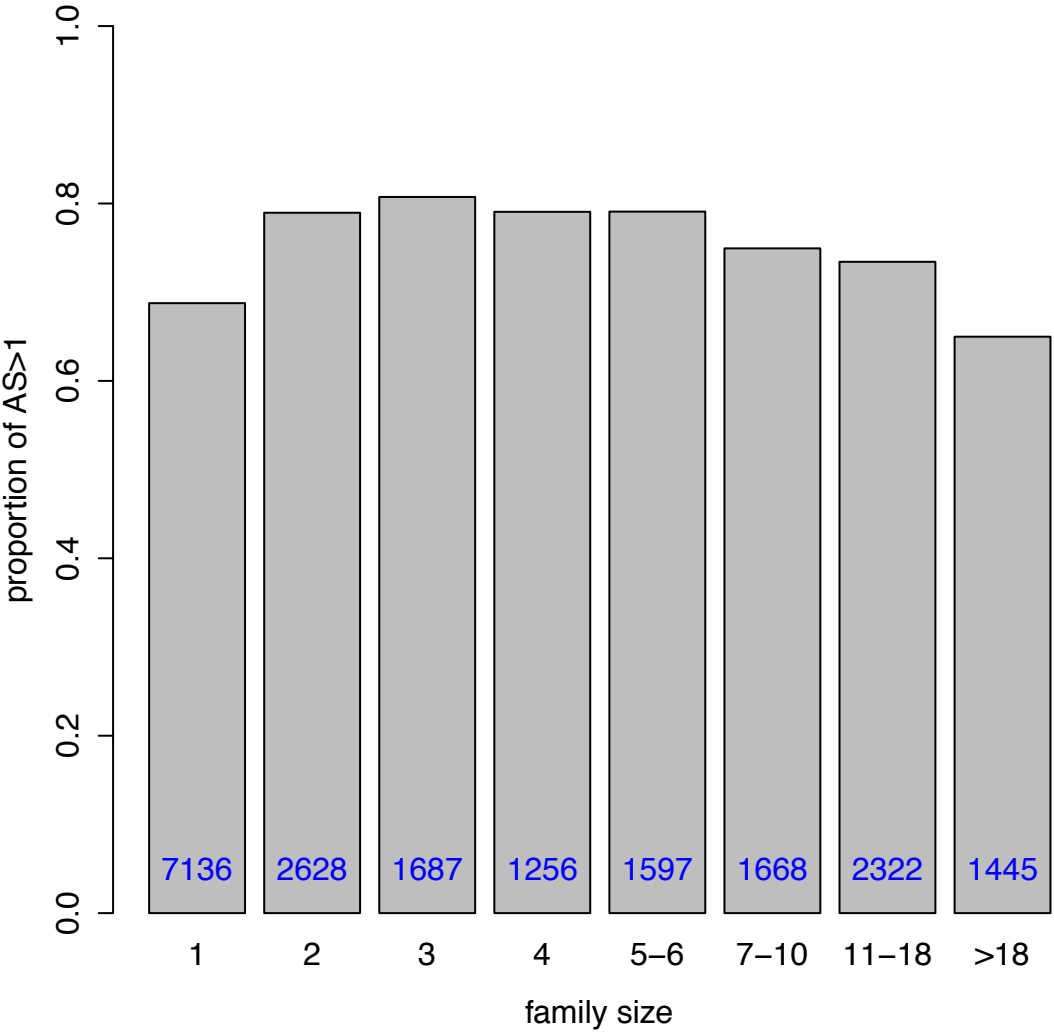

proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
7136
2628
1687
1256
1597
1668
2322
1445
1
2
3
4
5–6
7–10
11–18
>18
family size

Figure S12

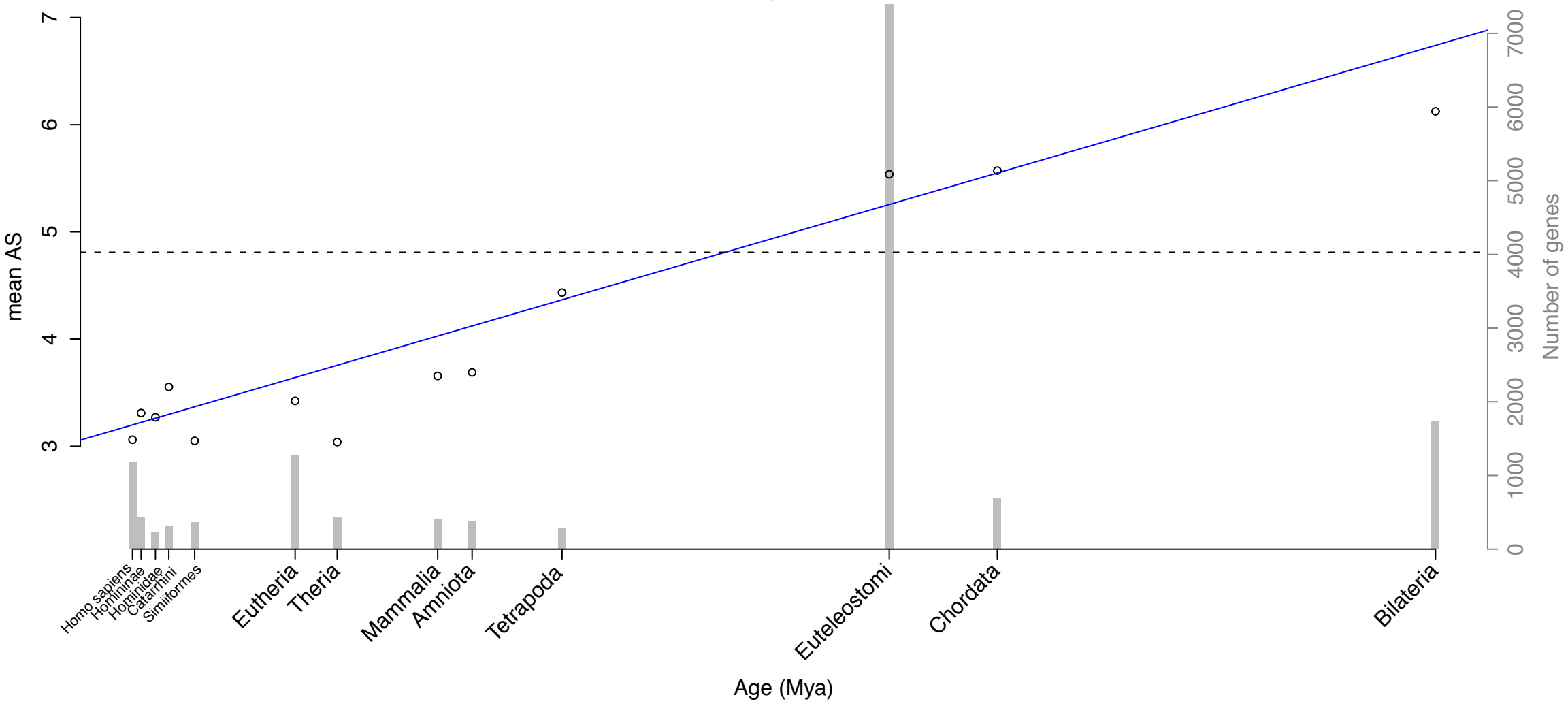

r = 0.95 / p = 4.9e−07
mean AS
Number of genes
Age (Mya)
Homo sapiens
Hominidae
Catarrhini
Simiiformes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria

Figure S13

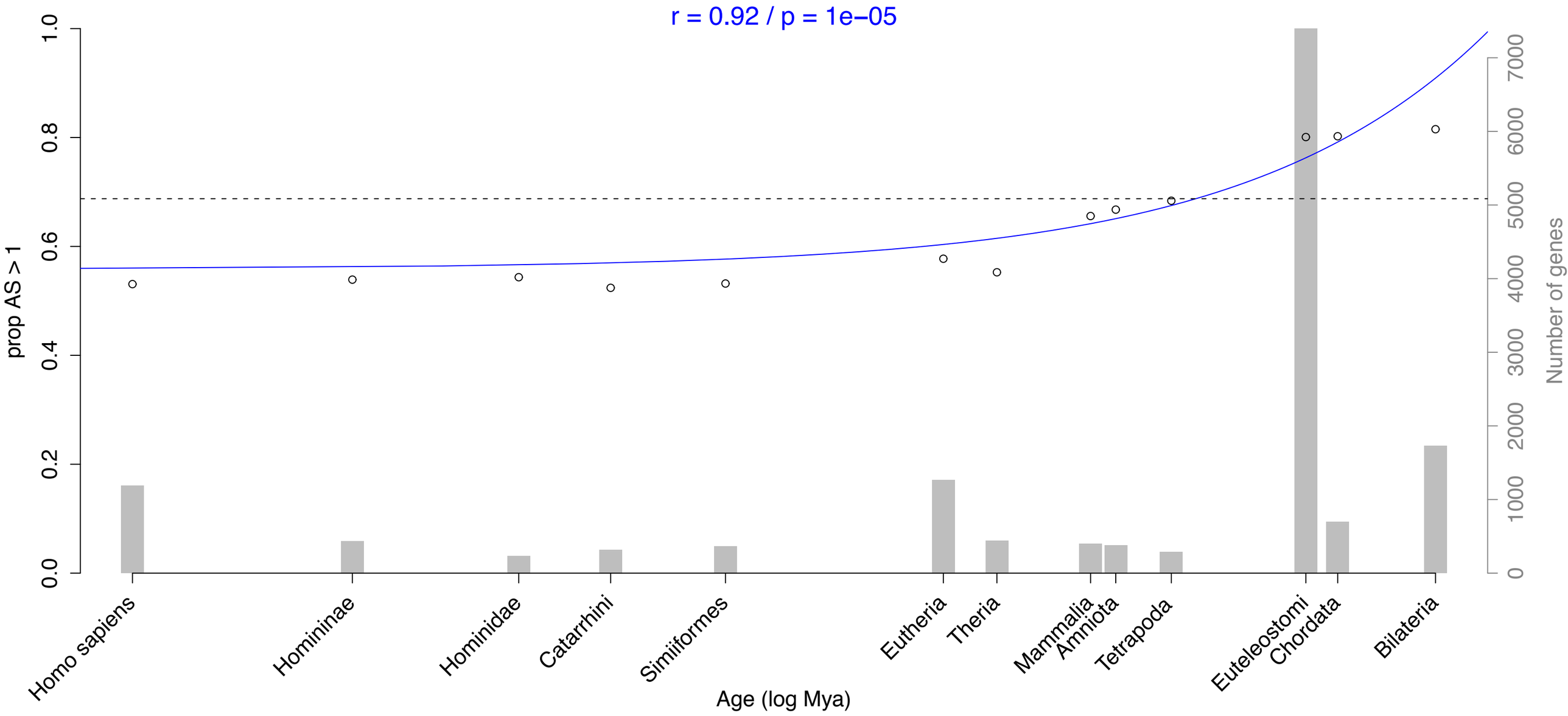
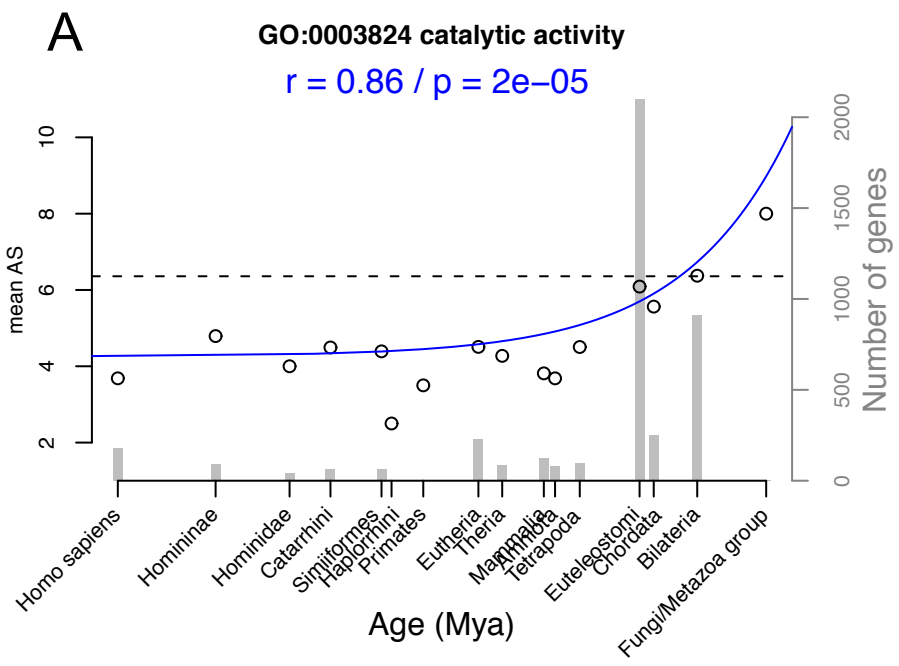
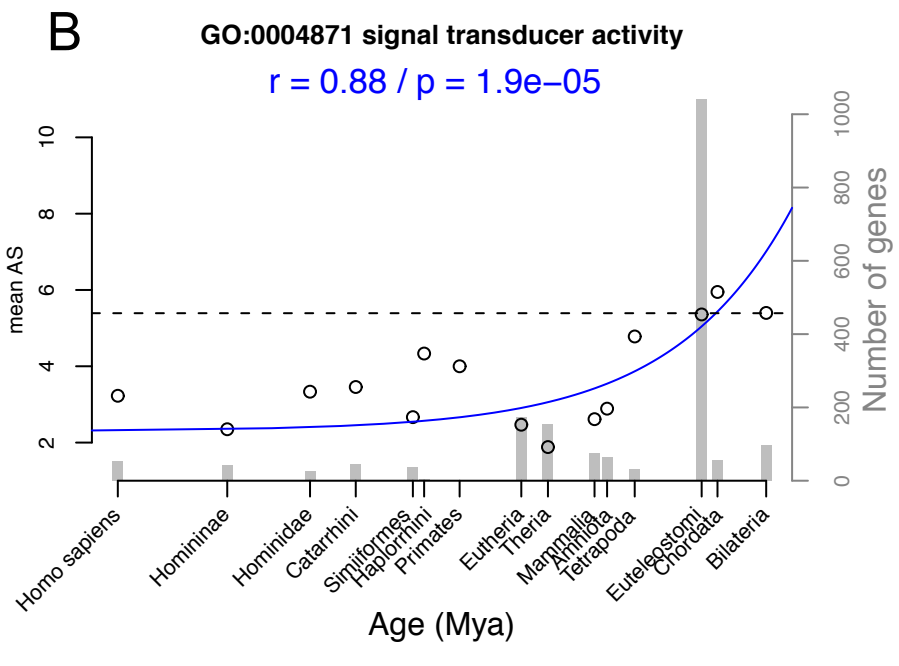
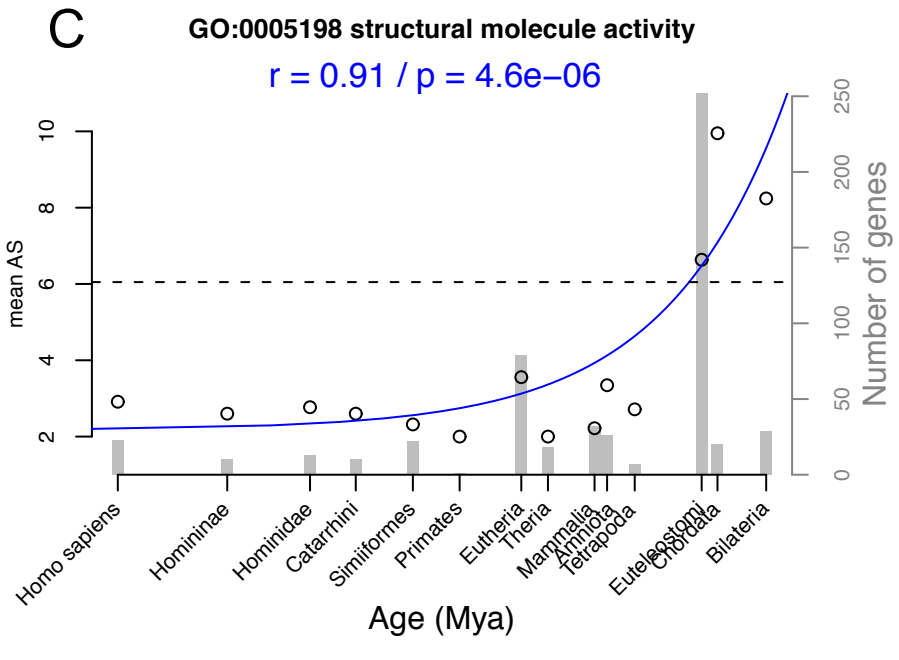
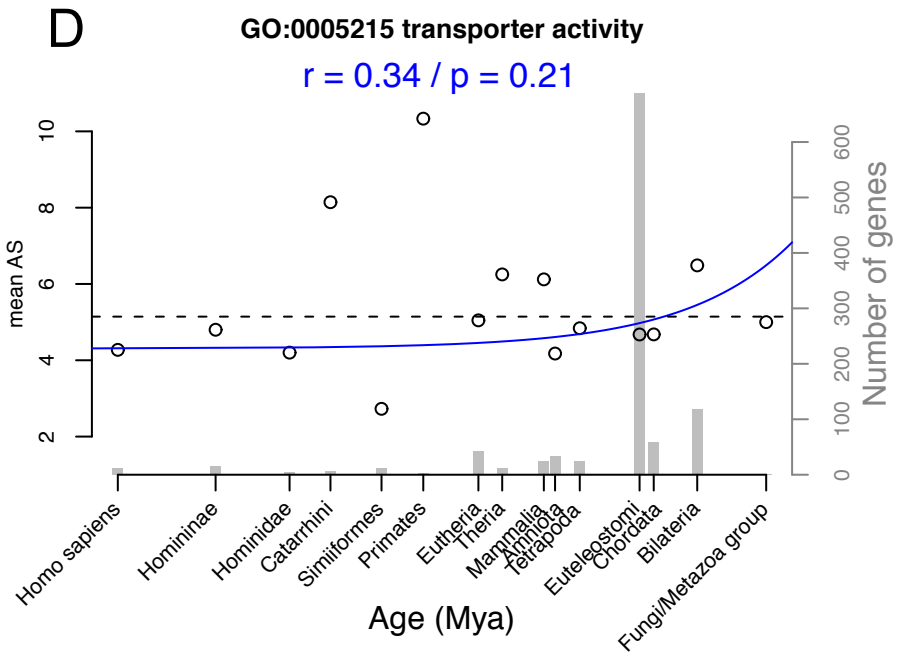

r = 0.92 / p = 1e−05
prop AS > 1
0.0
0.2
0.4
0.6
0.8
1.0
Number of genes
0
1000
2000
3000
4000
5000
6000
7000
Homo sapiens
Homininae
Hominidae
Catarrhini
Simiiformes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Age (log Mya)

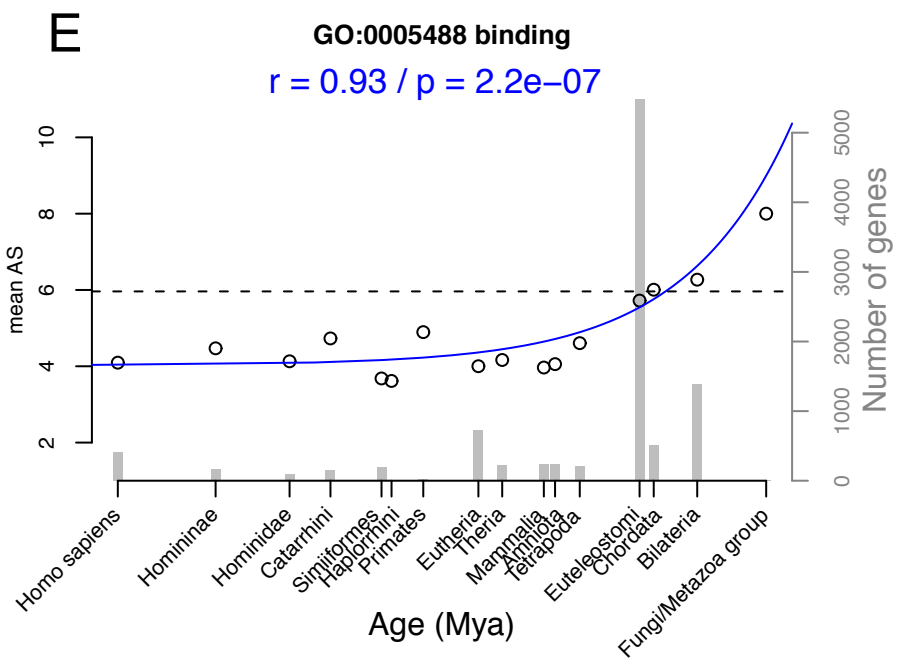
Figure S14

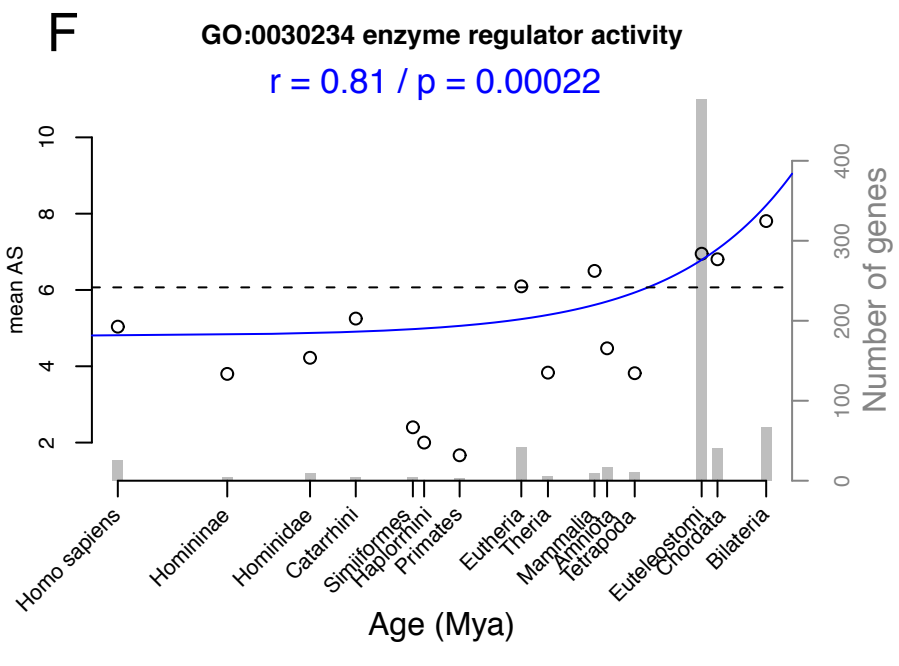

A
GO:0003824 catalytic activity
r = 0.86 / p = 2e−05
mean AS
Number of genes
Age (Mya)
Homo sapiens
Hominoidae
Hominidae
Catarrhini
Simiiformes
Haplorrhini
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Eutelegstomi
Chordata
Bilateria
Fungi/Metazoa group


B
GO:0004871 signal transducer activity
r = 0.88 / p = 1.9e−05
mean AS
Number of genes
Age (Mya)
Homo sapiens
Hominoidae
Hominidae
Catarrhini
Simiiformes
Haplorrhini
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Eutelegstomi
Chordata
Bilateria


C
GO:0005198 structural molecule activity
r = 0.91 / p = 4.6e−06
mean AS
Number of genes
Age (Mya)
Homo sapiens
Hominoidae
Hominidae
Catarrhini
Simiiformes
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Eutelegstomi
Chordata
Bilateria


D
GO:0005215 transporter activity
r = 0.34 / p = 0.21
mean AS
Number of genes
Age (Mya)
Homo sapiens
Hominoidae
Hominidae
Catarrhini
Simiiformes
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Eutelegstomi
Chordata
Bilateria
Fungi/Metazoa group


E
GO:0005488 binding
r = 0.93 / p = 2.2e−07
mean AS
Number of genes
Age (Mya)
Homo sapiens
Hominoidae
Hominidae
Catarrhini
Simiiformes
Haplorrhini
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Eutelegstomi
Chordata
Bilateria
Fungi/Metazoa group


F
GO:0030234 enzyme regulator activity
r = 0.81 / p = 0.00022
mean AS
Number of genes
Age (Mya)
Homo sapiens
Hominoidae
Hominidae
Catarrhini
Simiiformes
Haplorrhini
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Eutelegstomi
Chordata
Bilateria

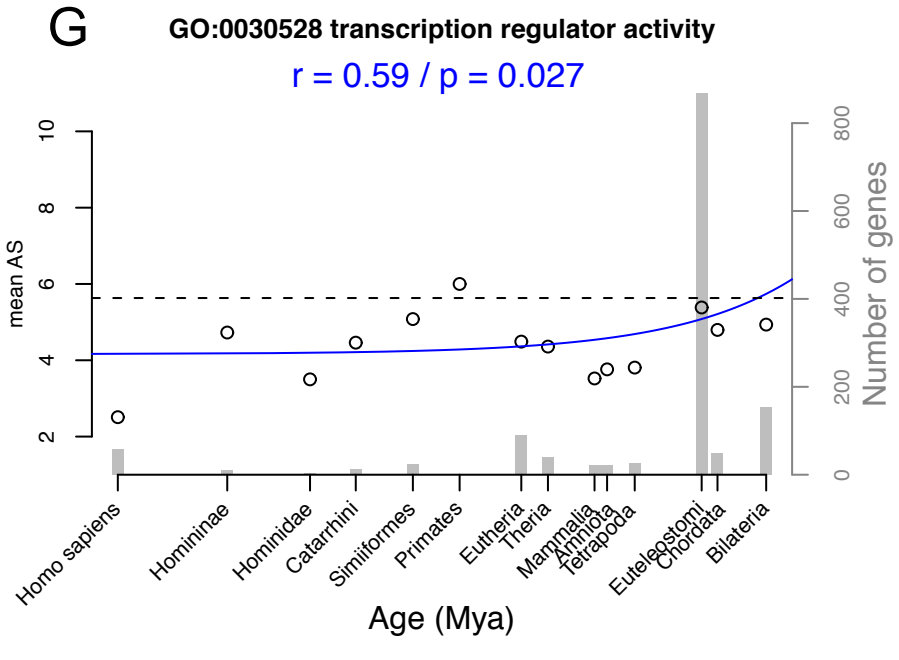

G
GO:0030528 transcription regulator activity
r = 0.59 / p = 0.027
mean AS
Number of genes
Age (Mya)
Homo sapiens
Hominoidae
Hominidae
Catarrhini
Simiiformes
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Eutelegstomi
Chordata
Bilateria

Figure S15

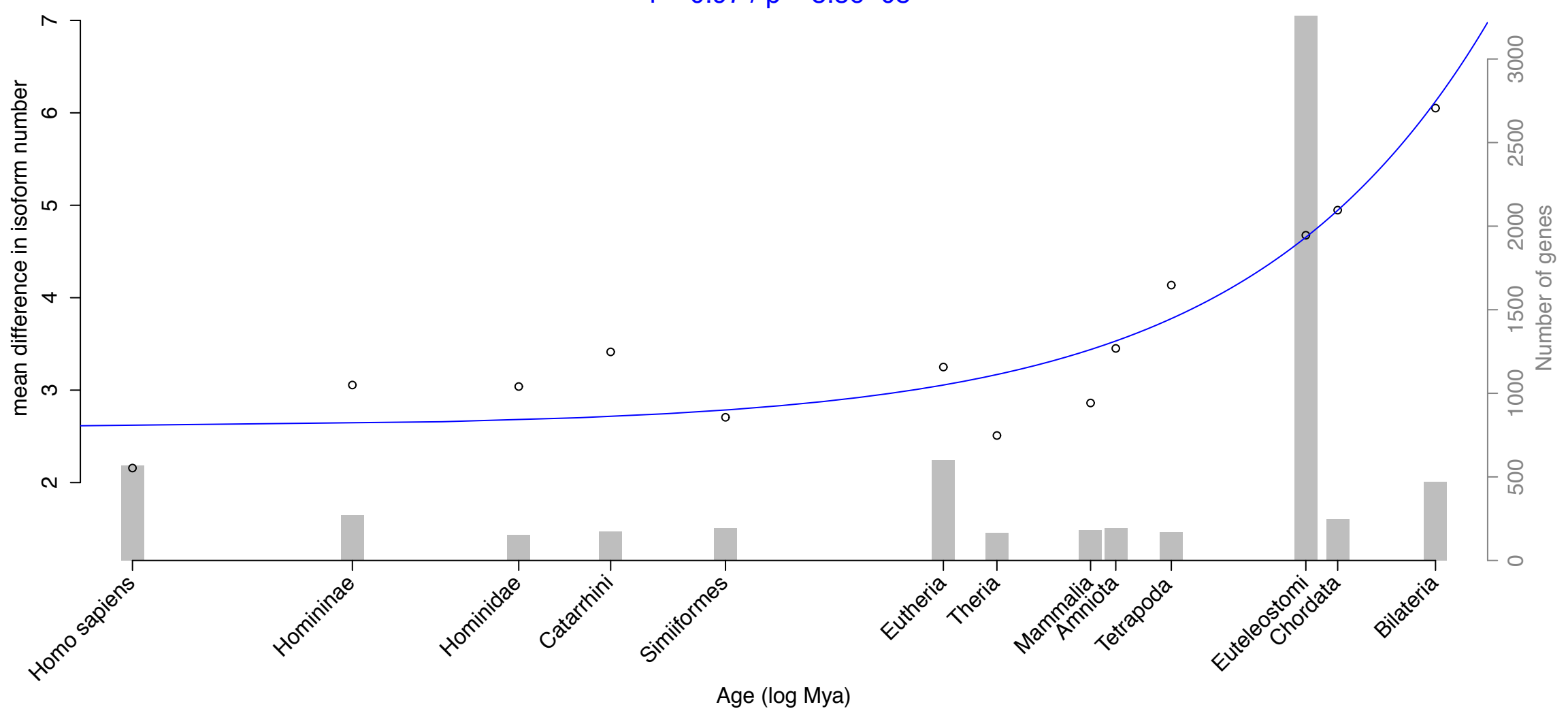

r = 0.97 / p = 5.5e−08
mean difference in isoform number
2
3
4
5
6
7
Number of genes
0
500
1000
1500
2000
2500
3000
Homo sapiens
Hominidae
Catarrhini
Simiiformes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Age (log Mya)

Figure S16

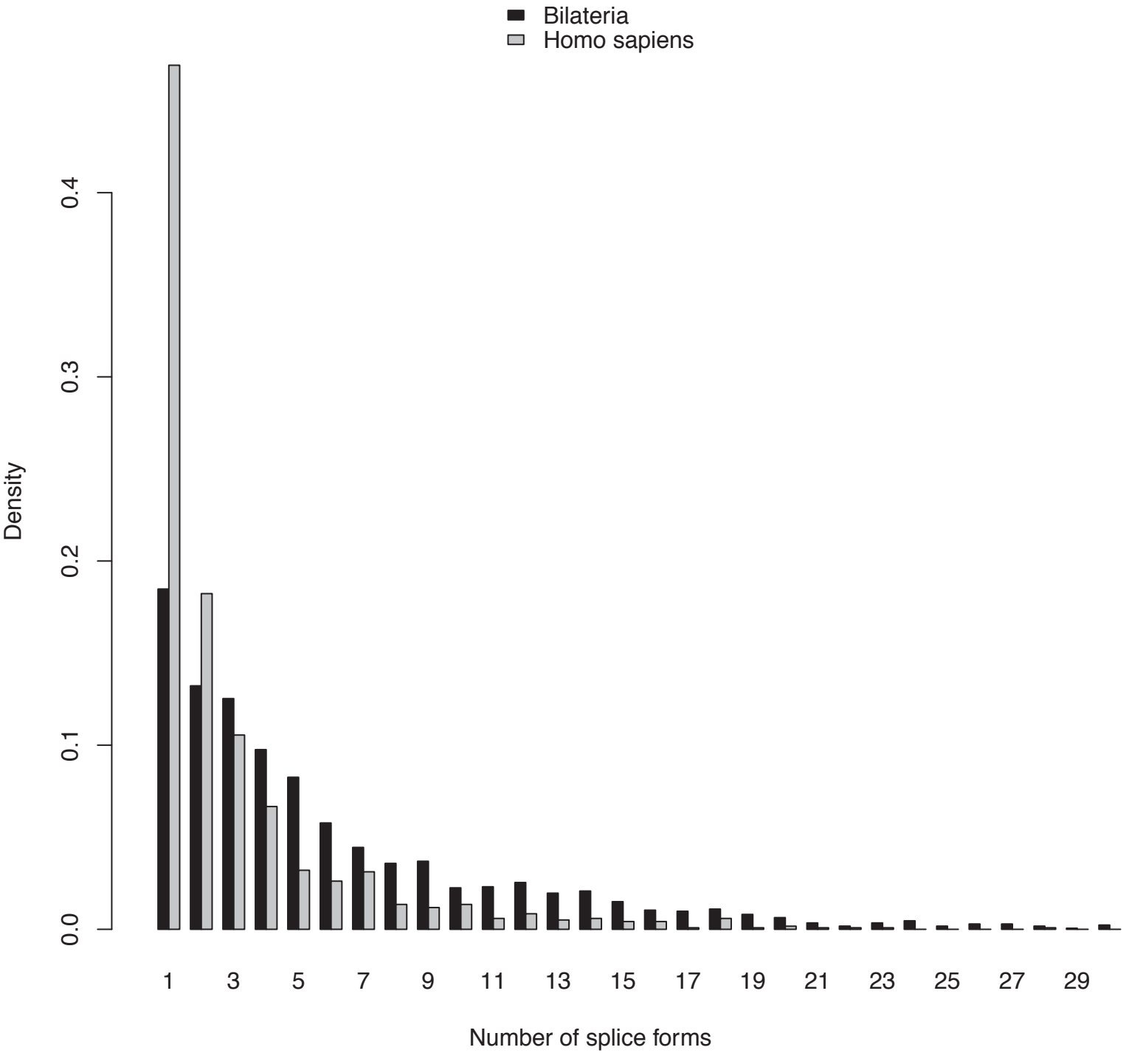

Bilateria
Homo sapiens
Density
0.0
0.1
0.2
0.3
0.4
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
Number of splice forms

Figure S17

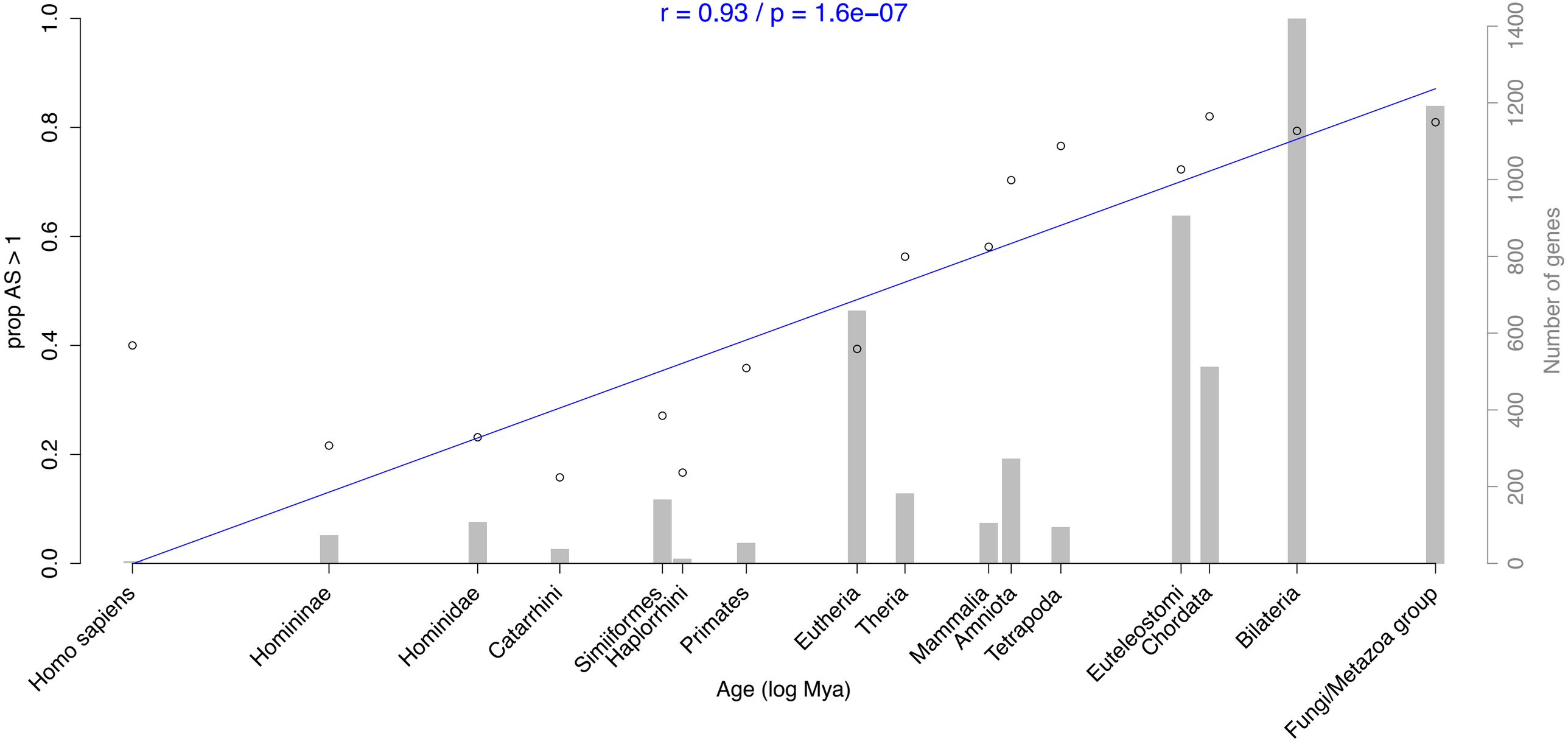

r = 0.93 / p = 1.6e−07
prop AS > 1
0.0
0.2
0.4
0.6
0.8
1.0
Number of genes
0
200
400
600
800
1000
1200
1400
Homo sapiens
Hominidae
Catarrhini
Simiiformes
Haplorrhini
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Fungi/Metazoa group
Age (log Mya)

Figure S18

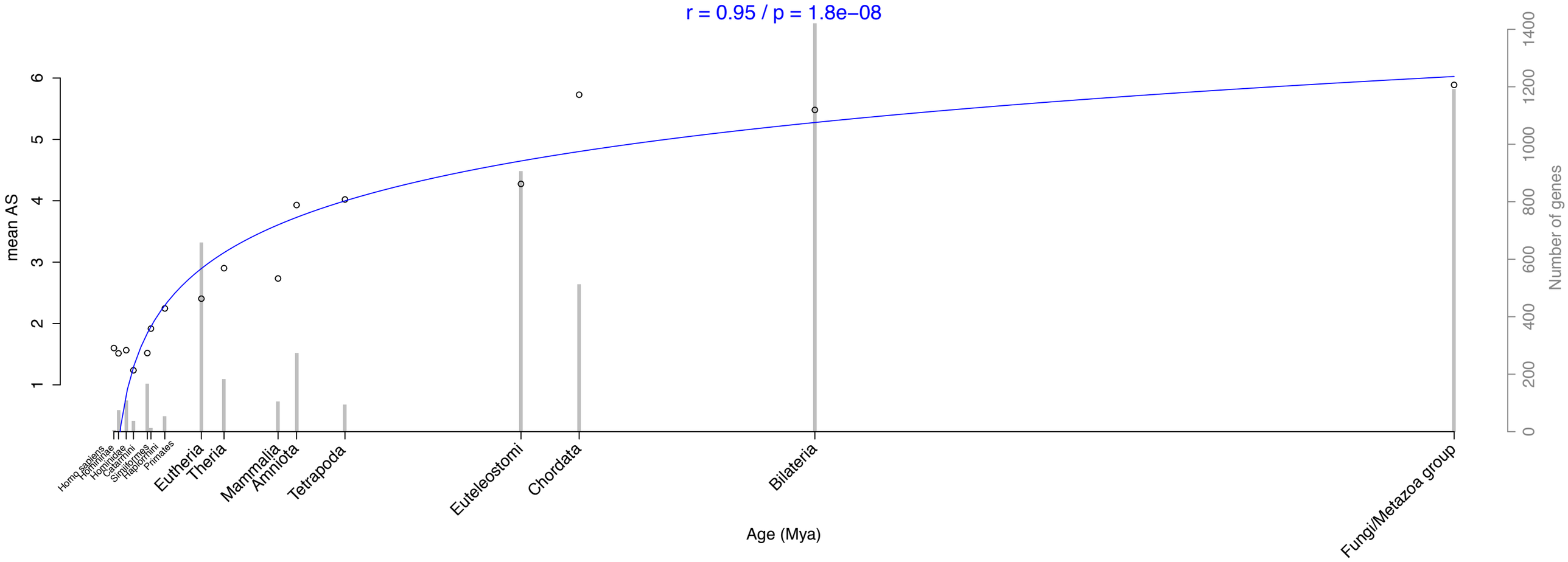
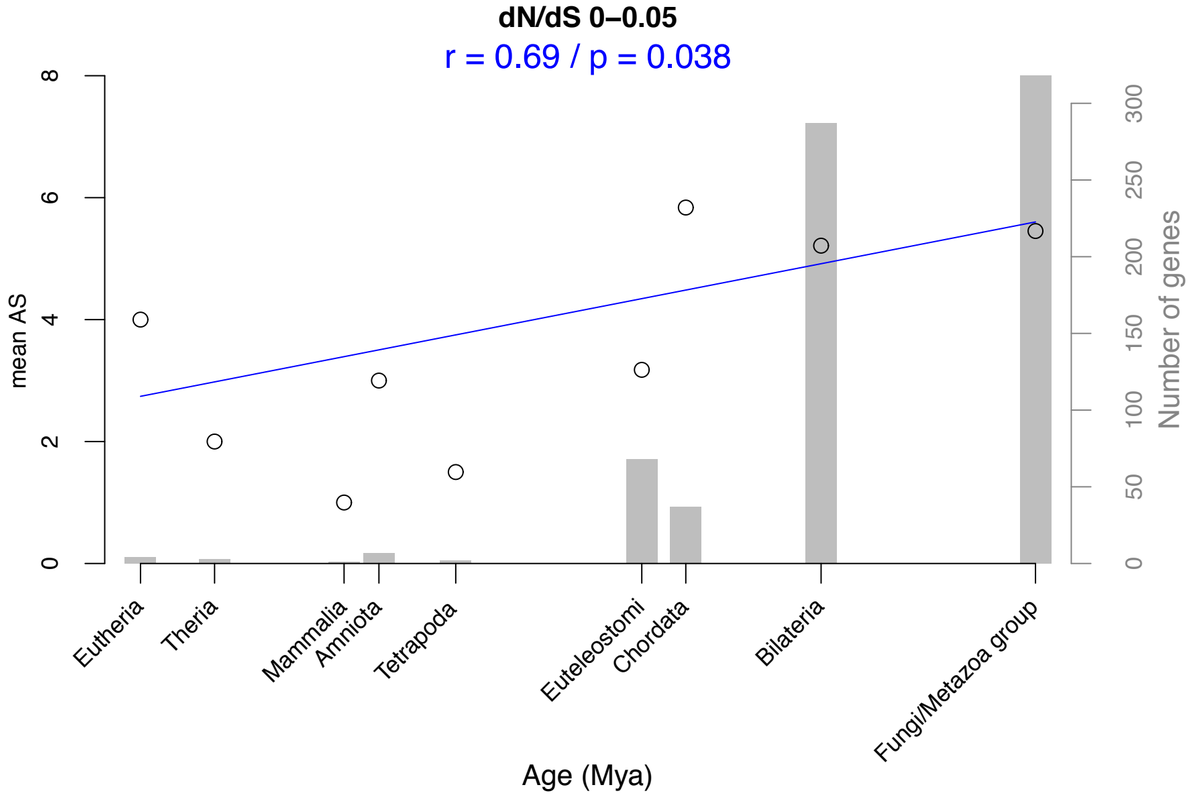
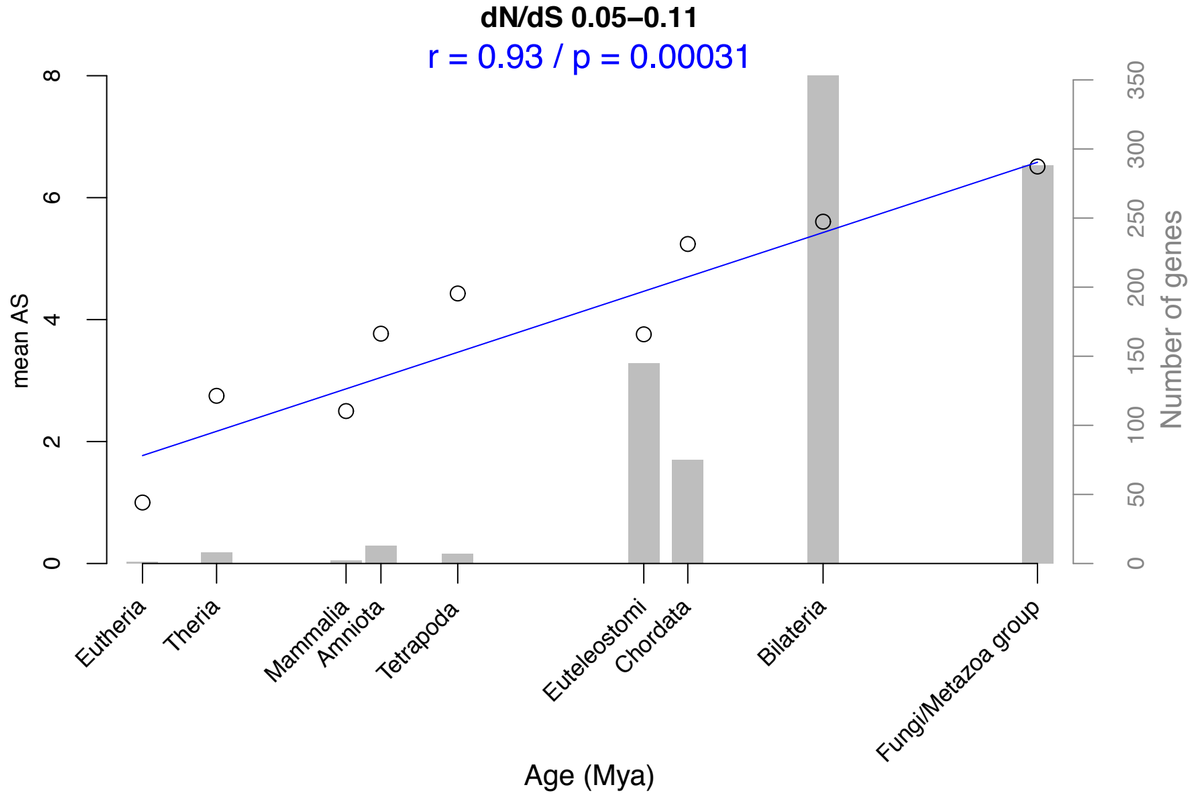
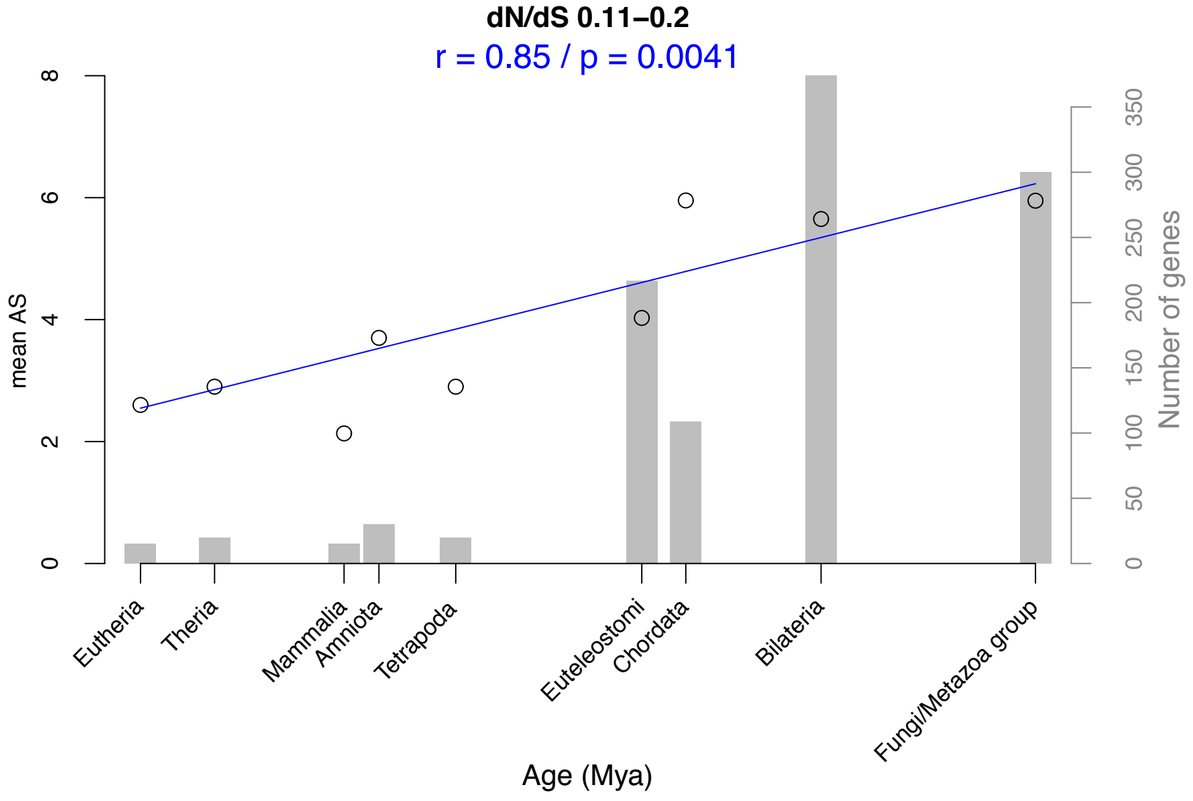

r = 0.95 / p = 1.8e−08
mean AS
1
2
3
4
5
6
Number of genes
0
200
400
600
800
1000
1200
1400
Homo sapiens
Homininae
Hominidae
Catarrhini
Simiiformes
Haplorrhini
Primates
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Fungi/Metazoa group
Age (Mya)

Figure S19


dN/dS 0–0.05
r = 0.69 / p = 0.038
mean AS
Number of genes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Fungi/Metazoa group
Age (Mya)


dN/dS 0.05–0.11
r = 0.93 / p = 0.00031
mean AS
Number of genes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Fungi/Metazoa group
Age (Mya)


dN/dS 0.11–0.2
r = 0.85 / p = 0.0041
mean AS
Number of genes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Fungi/Metazoa group
Age (Mya)

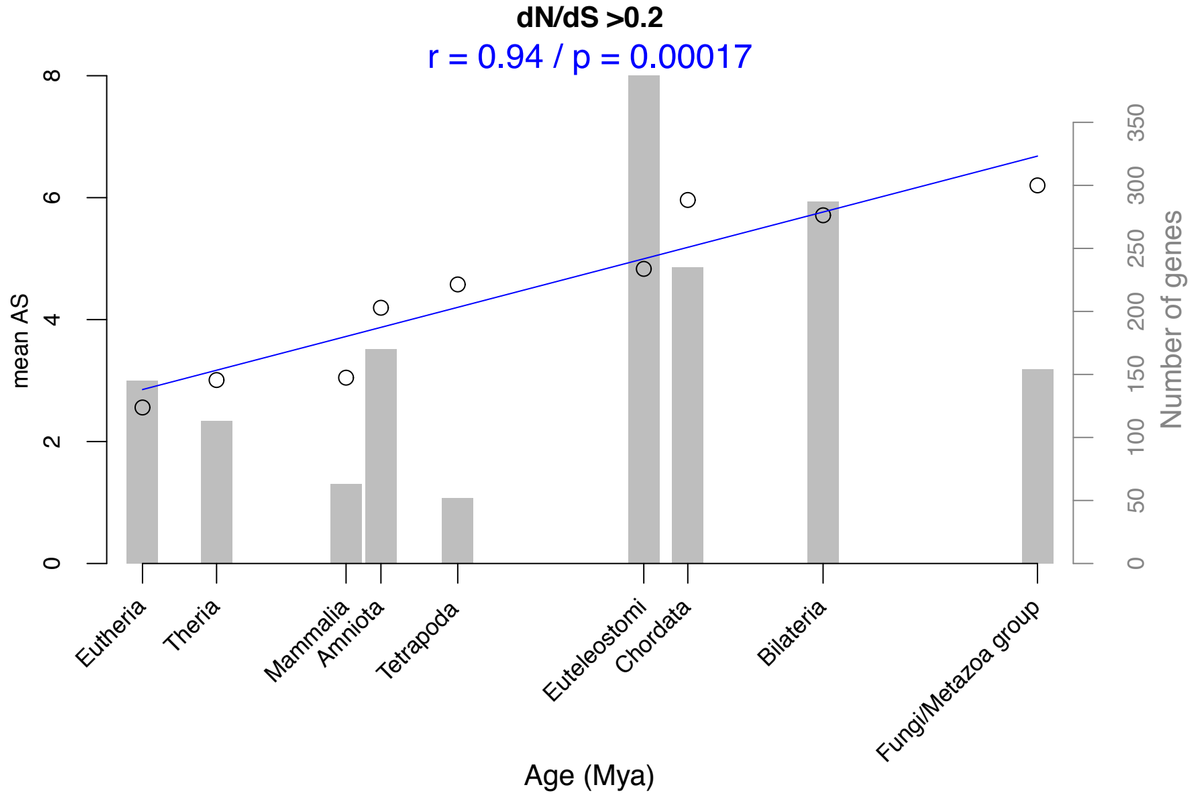

dN/dS >0.2
r = 0.94 / p = 0.00017
mean AS
Number of genes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Fungi/Metazoa group
Age (Mya)

Figure S20

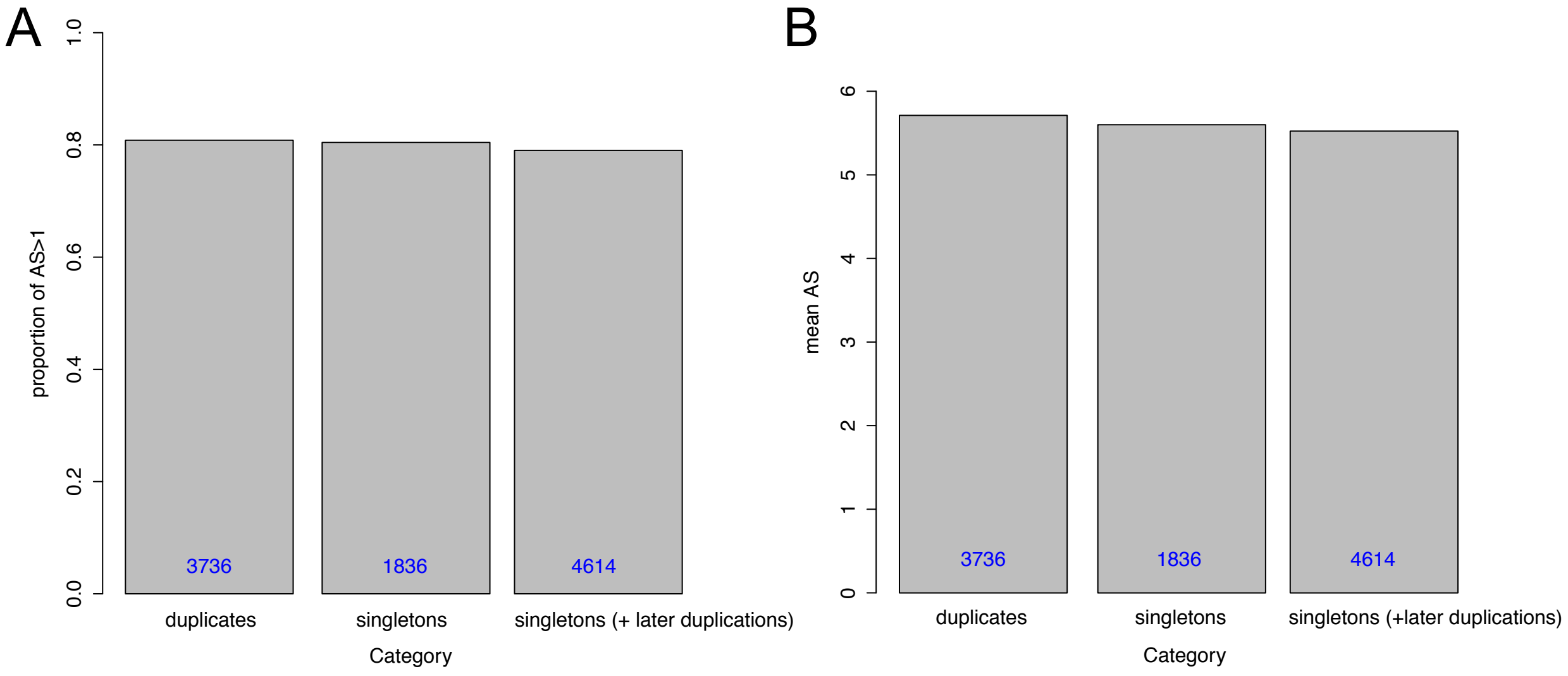
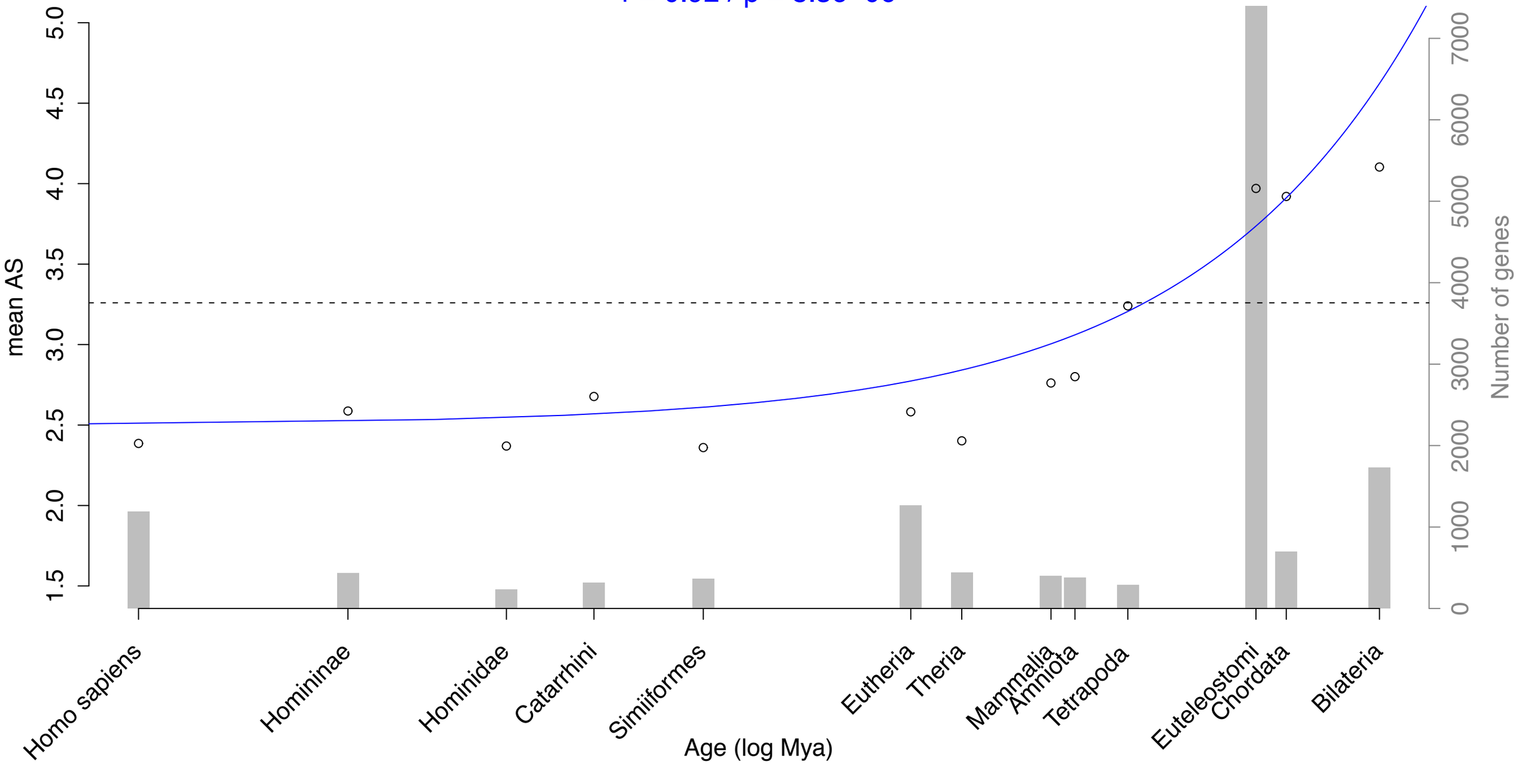

A
proportion of AS>1
0.0
0.2
0.4
0.6
0.8
1.0
3736
1836
4614
duplicates
singletons
singletons (+ later duplications)
Category
B
mean AS
0
1
2
3
4
5
6
3736
1836
4614
duplicates
singletons
singletons (+later duplications)
Category

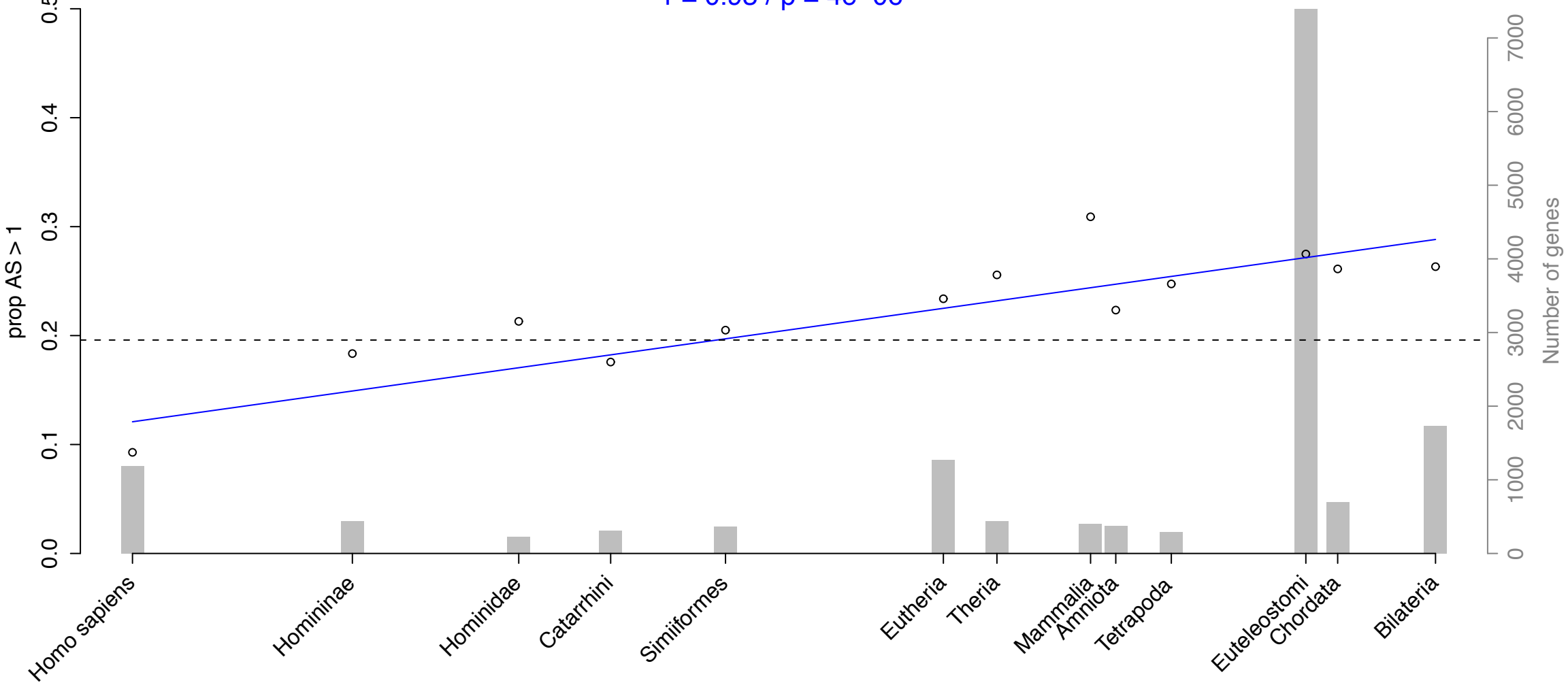
Figure S21


r = 0.92 / p = 8.3e−06
mean AS
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Number of genes
0
1000
2000
3000
4000
5000
6000
7000
Homo sapiens
Hominidae
Catarrhini
Simiiformes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Age (log Mya)

A

Figure S22

r = 0.93 / p = 4e−06



B

r = 0.5 / p = 0.083

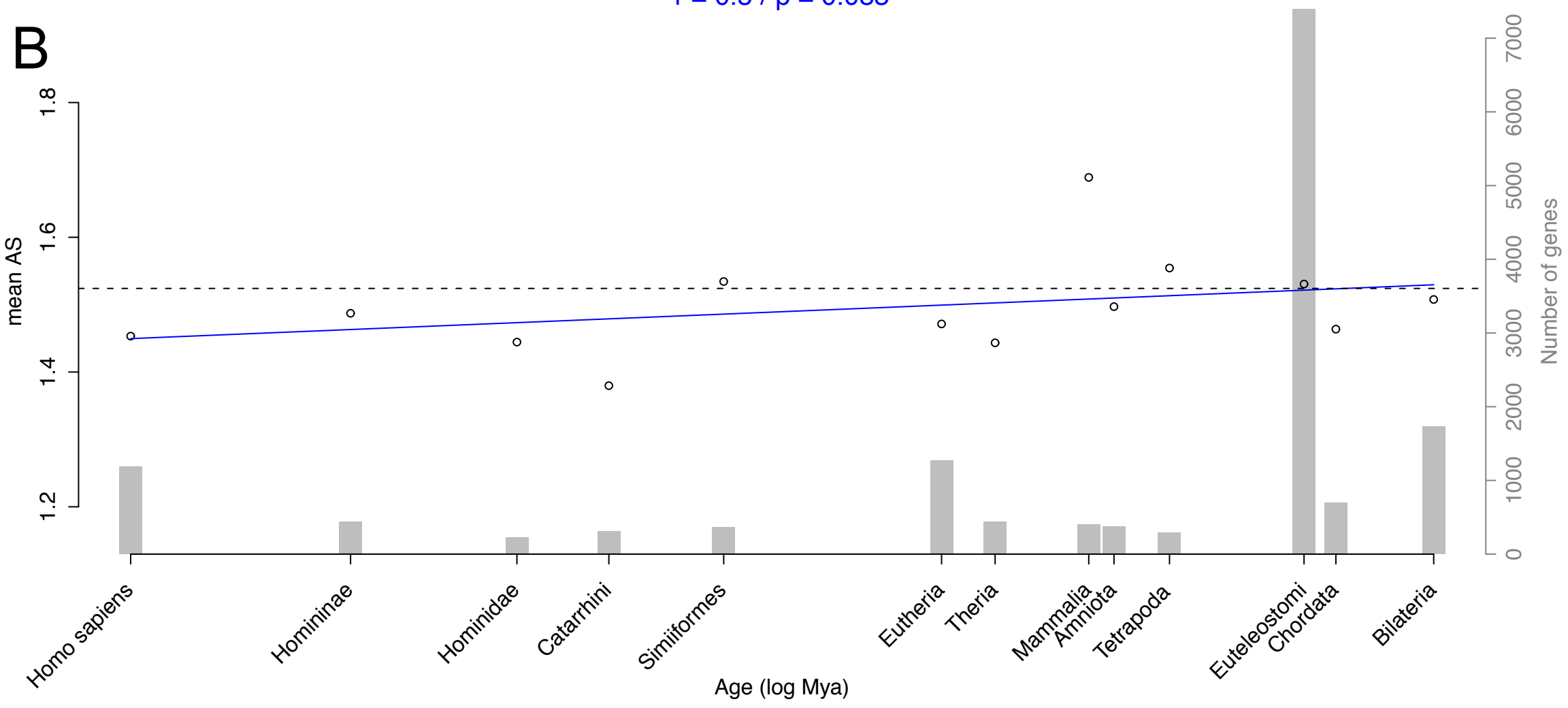


Figure S23

A
r = 0.95 / p = 4.8e−07
prop conserved AS > 1
0.00
0.05
0.10
0.15
0.20
Number of genes
0
1000
2000
3000
4000
5000
6000
7000
Homo sapiens
Hominidae
Catarrhini
Simiiformes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria

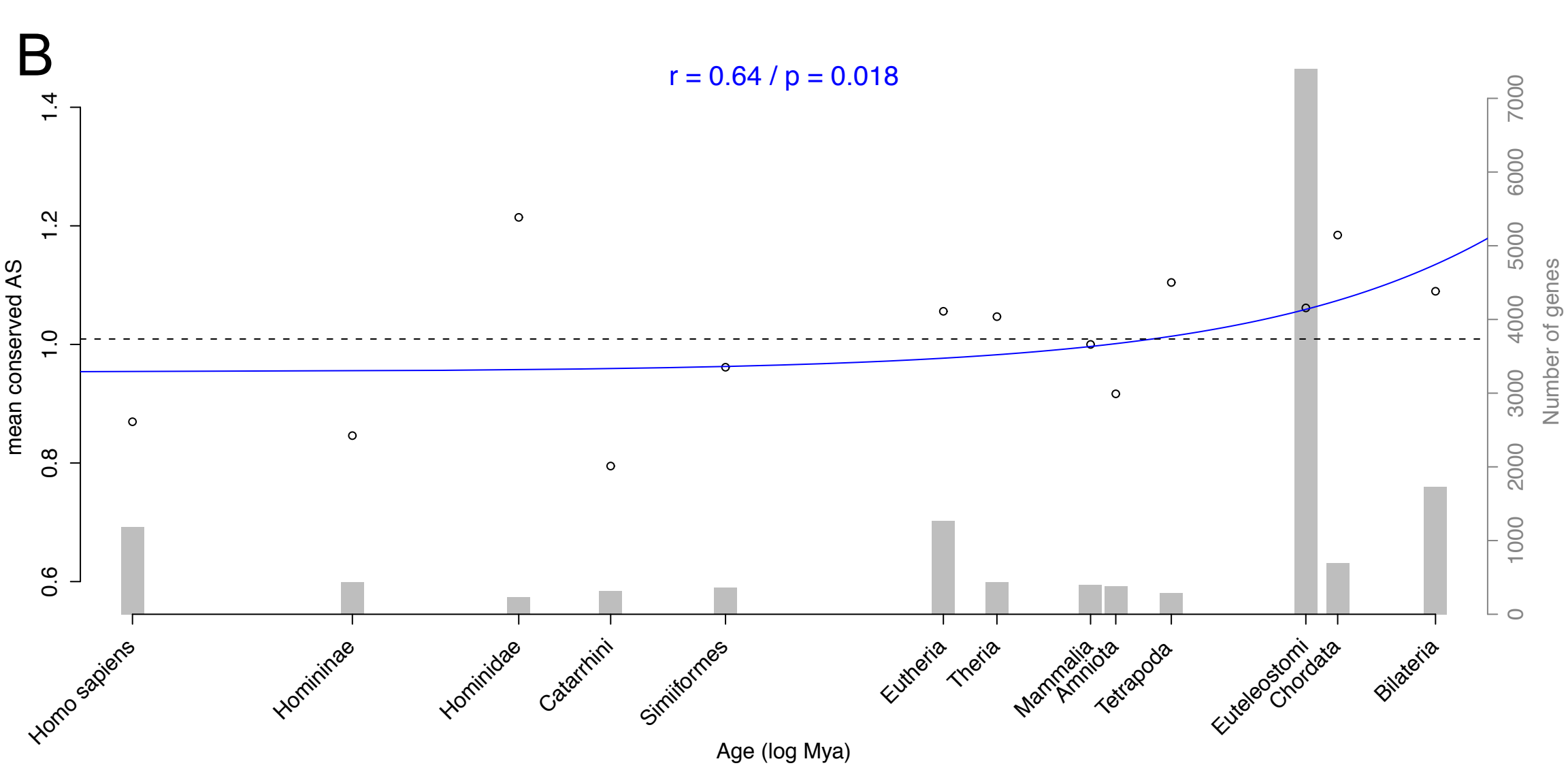

B
r = 0.64 / p = 0.018
mean conserved AS
0.6
0.8
1.0
1.2
1.4
Number of genes
0
1000
2000
3000
4000
5000
6000
7000
Homo sapiens
Hominidae
Catarrhini
Simiiformes
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Age (log Mya)

Figure S24

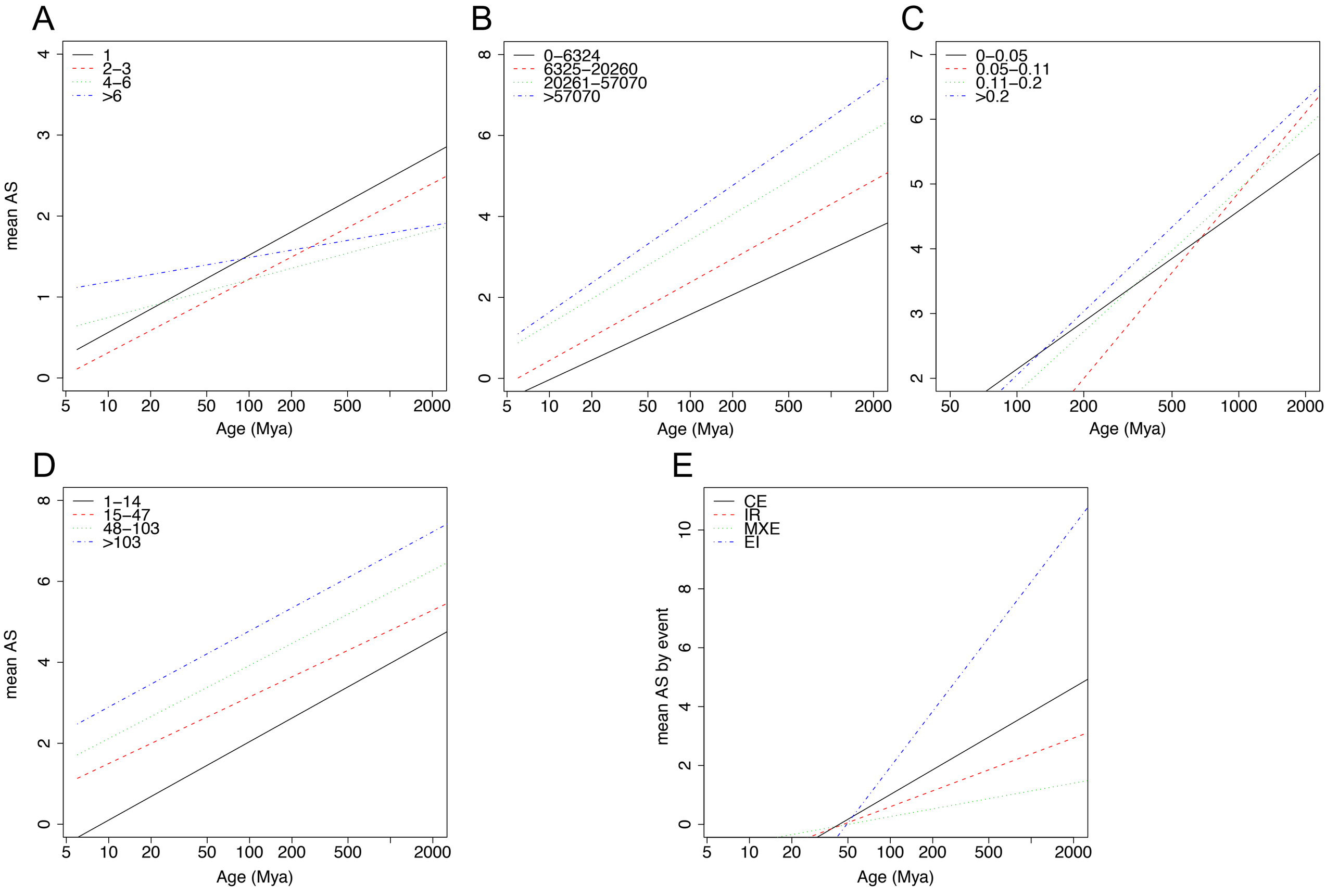
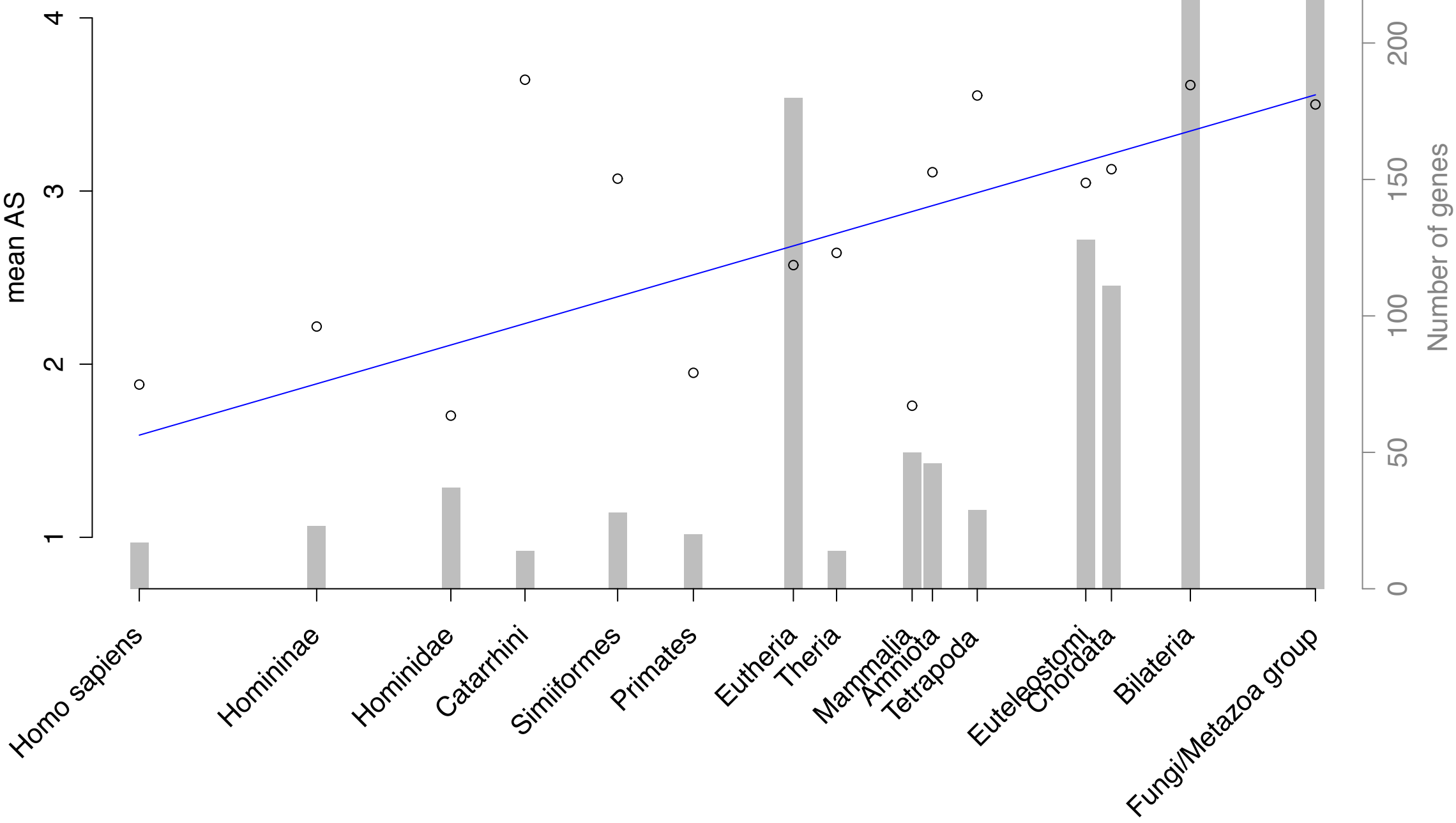

A
B
C
D
E
mean AS
1
2–3
4–6
>6
0–6324
6325–20260
20261–57070
>57070
0–0.05
0.05–0.11
0.11–0.2
>0.2
1–14
15–47
48–103
>103
mean AS by event
CE
IR
MXE
EI
Age (Mya)
5
10
20
50
100
200
500
1000
2000

Figure S25

A

r = 0.78 / p = 0.00059



B

r = 0.7 / p = 0.035

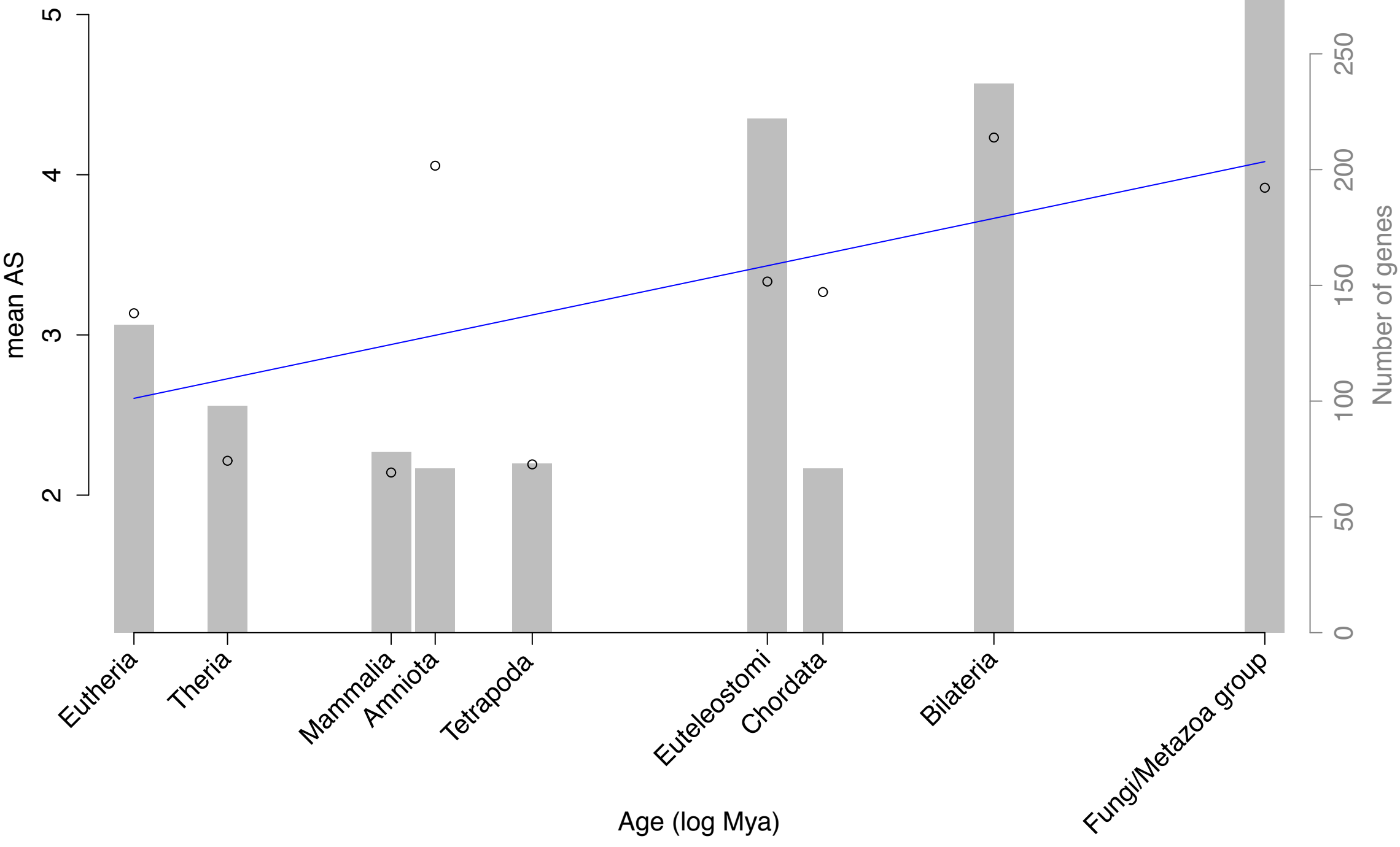


Figure S26

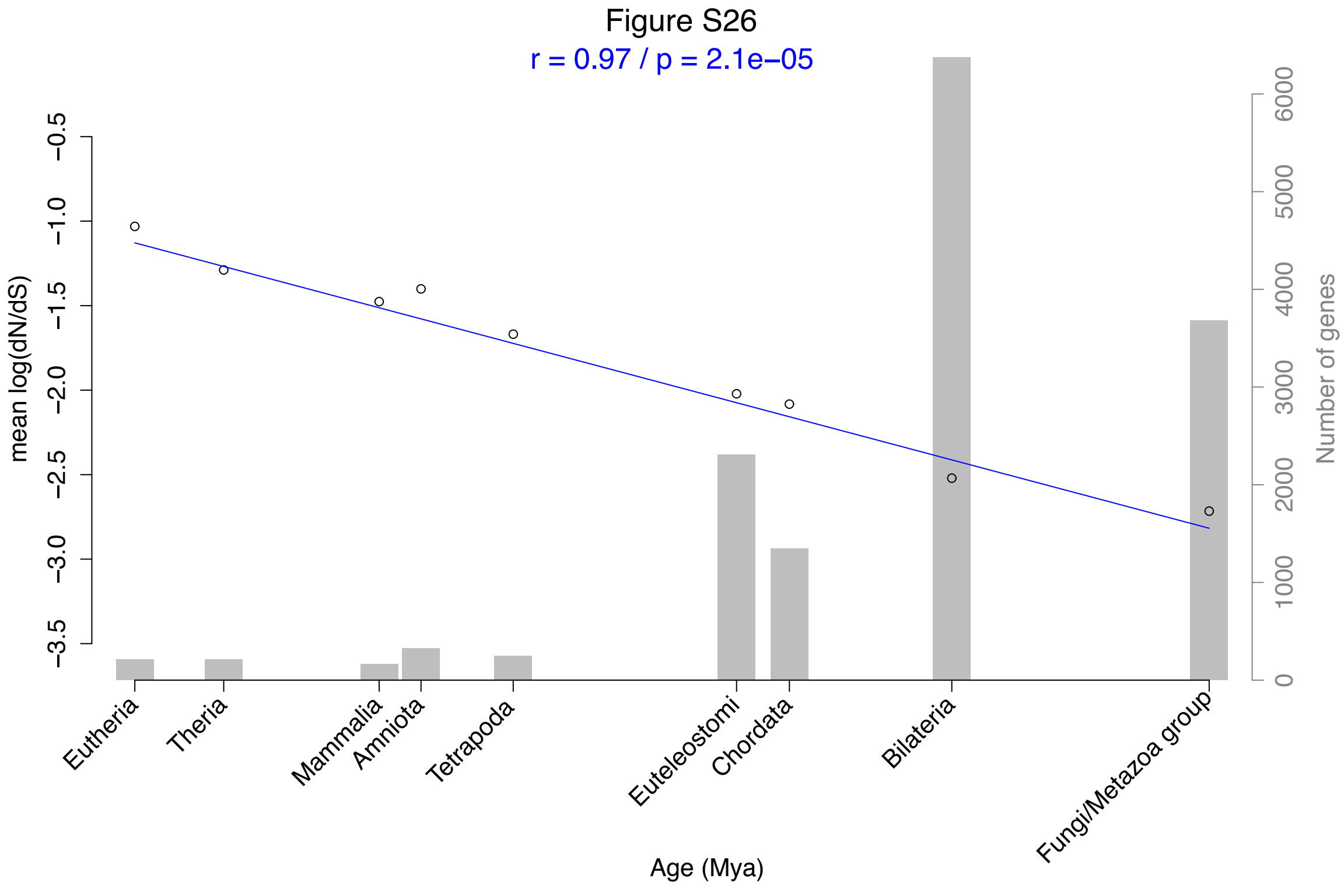
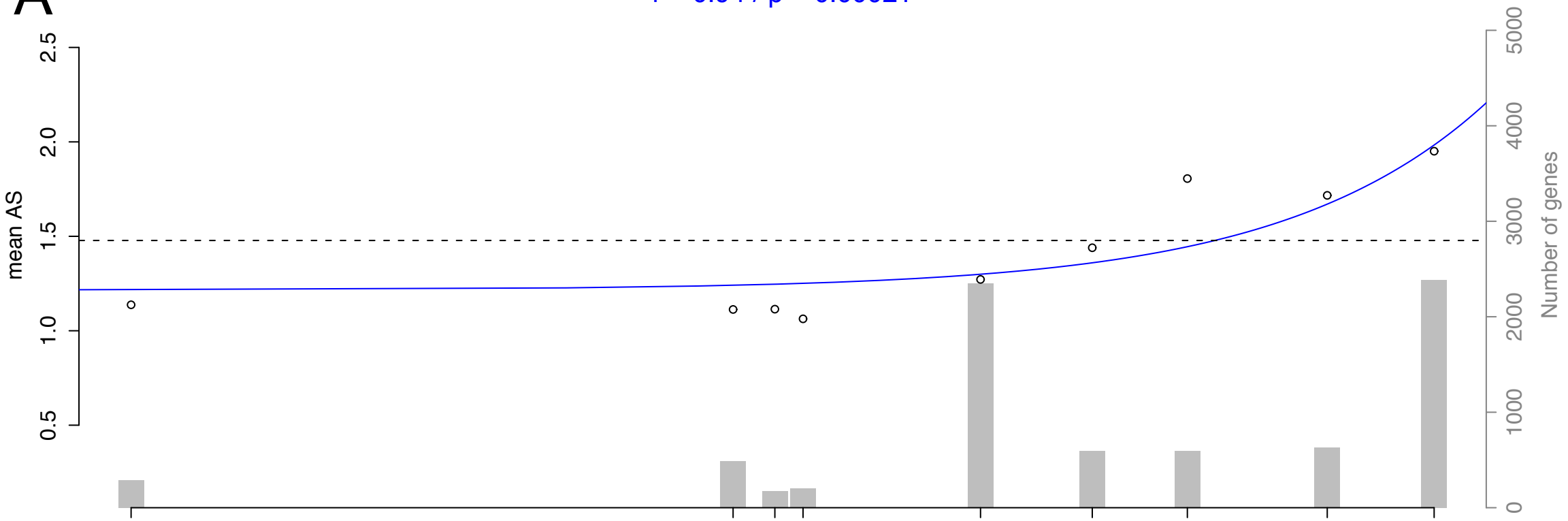

r = 0.97 / p = 2.1e−05
mean log(dN/dS)
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
Number of genes
0
1000
2000
3000
4000
5000
6000
Eutheria
Theria
Mammalia
Amniota
Tetrapoda
Euteleostomi
Chordata
Bilateria
Fungi/Metazoa group
Age (Mya)

Figure S27

A

r = 0.94 / p = 0.00021



B

r = 0.93 / p = 0.00034