

# Transposon- and genome dynamics in the fungal genus *Neurospora*: insights from nearly gapless genome assemblies

Diem Nguyen<sup>1</sup>, Valentina Peona<sup>1</sup>, Per Unneberg<sup>2</sup>, Alexander Suh<sup>1,3</sup>, Patric Jern<sup>4</sup>, Hanna Johannesson<sup>1\*</sup>

<sup>1</sup> Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, SE-75236 Uppsala, Sweden

<sup>2</sup> Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, SE-75237 Uppsala, Sweden

<sup>3</sup> School of Biological Sciences, University of East Anglia, Norwich Research Park, NR4 7TU, Norwich, United Kingdom

<sup>4</sup> Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, SE-75123 Uppsala, Sweden

\*Corresponding author: H. Johannesson, hanna.johannesson@ebc.uu.se

## Abstract

A large portion of nuclear DNA is composed of transposable element (TE) sequences, whose transposition is controlled by diverse host defense strategies in order to maintain genomic integrity. One such strategy is the fungal-specific Repeat-Induced Point mutation (RIP) that hyper-mutates repetitive DNA sequences. While RIP is found across Fungi, it has been shown to vary in efficiency. The filamentous ascomycete *Neurospora crassa* has been a pioneer in the study of RIP, but data on TEs and RIP from other species in the genus is limited. In this study, we investigated 18 nearly gapless genome assemblies of ten *Neurospora* species, which diverged from a common ancestor about 7 MYA, to determine and compare genome-wide TE distribution and their associated RIP patterns. Four of these assemblies, generated by PacBio technology, represent new genomic datasets. We showed that the TE contents (8.7-18.9%) covary with genome sizes that range between 37.8-43.9 Mb. Degraded copies of Long Terminal Repeat (LTR) retrotransposons were abundant among the identified TEs, and these are distributed across the genome at varying frequencies. In all investigated *Neurospora* genomes, TE sequences had signs of numerous C-to-T substitutions, suggesting that RIP occurred in all species, and accordingly, RIP signatures correlated with TE-dense regions in all genomes. In conclusion, essentially gapless genome assemblies allowed us to identify TEs in *Neurospora* genomes, and reveal that TEs contribute to genome size variation in this group. Our study suggests that TEs and RIP are highly correlated in each examined *Neurospora* species, and hence, the pattern of interaction is conserved over the investigated evolutionary timescale. Finally, with our results, we verify that RIP signatures can be used to facilitate the identification of TE-rich regions in the genome. The comprehensive genomic dataset of *Neurospora* is a rich resource for further in-depth analyses of fungal genomes by the community.

## Introduction

Advanced sequencing technologies have yielded numerous genome assemblies of high quality from diverse species across the tree of life (O'leary *et al.* 2015). Despite the availability of these assemblies, there are still difficult-to-sequence genomic regions, such as repetitive sequences and centromeres (Johnson *et al.* 2005; Peona *et al.* 2018). The difficulty to sequence these regions has limited studies of genome architecture and evolution, which depend on data from multiple

genomes. Thus, availability of high-quality gapless genome assemblies from multiple individuals and species are critically needed, particularly to understand how repetitive sequences contribute to genome size evolution and how genome integrity is maintained.

Genomes are dynamic entities, subject to large and small-scale rearrangement events that can lead to gains and losses of genomic DNA sequences. Transposable elements (TEs) are genetic elements present in eukaryotic and prokaryotic genomes, characterized by their ability to propagate in the genome. While TEs may contribute to evolutionary change and innovation (Werren 2011), TE activity can lead to deleterious effects including insertional disruption of functionally important sequences such as promoters or genes themselves, gene miss-expression, and silencing of adjacent genes (Hollister and Gaut 2009; Chuong *et al.* 2017). TEs can also mediate ectopic recombination between distant chromosomal regions, inversions, and deletion of genomic sequences (Cáceres *et al.* 1999; Gray 2000).

Based on the nature of their transposition intermediates, TEs can be classified as Class I RNA retrotransposons (transposition via RNA intermediate) or Class II DNA transposons (transposition via DNA intermediate) (Wicker *et al.* 2007; Makiłowski *et al.* 2019). They are further divided into subclasses, orders, and super-families based on their specific structural and coding features. Class I elements, consisting of LTR retroelements, LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements), are the most frequent class of TEs in many animals (e.g., *Drosophila melanogaster*), plants (e.g., *Arabidopsis thaliana* and maize), and fungi (Arabidopsis Genome 2000; Feschotte and Pritham 2007; Schnable *et al.* 2009; Amselem *et al.* 2015).

To counterbalance potential detrimental side effects of genome-damaging agents, genome defense mechanisms have evolved to ensure the maintenance of genome integrity (Matzke *et al.* 2000; Selker 2004; Johnson 2007; Blumenstiel 2011). Eukaryotes have evolved ways to defend their genomes against TEs (Matzke *et al.* 2000; Gladyshev and Kleckner 2016). For example, plants and animals utilize methylation and RNA interference (RNAi)-based mechanisms (Colot and Rossignol 1999; Zeng *et al.* 2019). In fungi, one such mechanism is sequence homology-based Repeat-Induced Point mutation (RIP) (Selker 1990; Selker 2002). First discovered in *Neurospora crassa*, RIP permanently mutates duplicated sequences, such as TEs, and introduces multiple C-to-T transition mutations, predominantly at CpA dinucleotides, in both copies of the sequences, skewing dinucleotide frequencies towards an over-representation of TpA in mutated sequences (Selker 1990; Selker 2002). For RIP to occur, there are requirements for minimal duplicated sequence length of about 400 base pairs (bp), though sequences as short as 155 bp have been reported, and sequence identity of greater than 80% (Cambareri *et al.* 1991; Watters *et al.* 1999; Gladyshev and Kleckner 2014). In evolutionarily diverged fungal species where RIP has been experimentally demonstrated, intact and active TEs are still present in the genomes (reviewed in (Hane *et al.* 2015)). In contrast, RIP is considered highly efficient in *N. crassa* (but see (Singer *et al.* 1995)) and has resulted in the essential absence of intact and active TEs (Galagan and Selker 2004).

Here we study the distribution of TEs and RIP in the filamentous fungal genus *Neurospora*. The motivation of our study is fourfold. First, the genus *Neurospora* represents a tractable system; it has small genomes relative to other eukaryotes (~43 Mb) (Galagan *et al.* 2003; Borkovich *et al.* 2004), and the model organism *N. crassa* has a well-annotated genome (Galazka *et al.* 2016). Second, in addition to the well-studied *N. crassa*, the genus contains multiple species for which the species boundaries and phylogenetic relationship is well resolved (Dettman *et al.* 2003; Villalta *et al.* 2009; Corcoran *et al.* 2012; Corcoran *et al.* 2014). The last

common ancestor of the *Neurospora* clade diverged within the last 7 million years, hence the system provides a framework for comparative studies over relatively short evolutionary timescales (Whittle and Johannesson 2012). Third, despite its relatively small genome, about 10% of the *Neurospora* genome is comprised of repetitive DNA sequences (Galagan *et al.* 2003; Hosseini *et al.* 2020). Fourth, the efficiency of RIP varies across the kingdom Fungi (Galagan and Selker 2004; Hane *et al.* 2015). Previous studies on the conservation of the gene *rid*, which is essential for RIP, in *N. tetrasperma* and *N. intermedia* (Freitag *et al.* 2002) and on the mutational bias in *N. tetrasperma* (Ellison *et al.* 2011), have indicated conservation of the RIP-process in *Neurospora*, but no systematic survey of the TE and RIP pattern in the genus have been carried out so far. Altogether, *Neurospora* genomes present a well-suited model system for studying the relationship between TEs and genome defense.

Our study of the TE landscape in *Neurospora* was facilitated by high-quality nearly gapless genome assemblies of long-read sequenced genomes. We thus comprehensively surveyed TEs based on their diversity, sequence abundance, and chromosomal distribution in each genome assembly, and compared these features between and within species to better describe the dynamics of TE variability. Furthermore, we determined how RIP associated with TEs among the different genomes in this *Neurospora* clade to better understand whether RIP efficiency in *N. crassa* is a general feature of *Neurospora*. Finally, with our study, we confirm the application of genome-wide RIP index to identify TE enriched locations in the genome, when identifying TEs in genomes heavily affected by RIP is not possible.

## Results

### *Nearly gapless Neurospora assemblies*

To study the TE and RIP landscapes of closely-related *Neurospora* species (Figure 1) (Menkis *et al.* 2009; Corcoran *et al.* 2014; Corcoran *et al.* 2016), we first generated PacBio assemblies from four genomes, providing new genomic datasets. Two of these genomes previously lacked genomic information [*N. hispaniola* (FGSC 8817), and *N. metzenbergii* (FGSC 10397)] (Table S1), and from *N. intermedia* (FGSC 8767) and *N. discreta* (FGSC 8579), genomic data generated using short-read sequencing technologies have previously been published (Grigoriev *et al.* 2011; Svedberg *et al.* 2018). Additionally, we used 11 previously published PacBio-sequenced genome assemblies [i.e. from the genomes of *N. intermedia* (FGSC 8761, 8807), *N. sitophila* (FGSC 5940, 5941, W1426, and W1434), and *N. tetrasperma* (FGSC 2503, 2504, 9045, 9046, and 10752)] (Jacobson *et al.* 2006; Sun *et al.* 2017; Svedberg *et al.* 2018; Hosseini *et al.* 2020; Svedberg *et al.* 2021). Reference assemblies are available for *N. crassa* (FGSC 2489) and two strains of *N. tetrasperma* L6 (FGSC 2508, 2509) (Galagan *et al.* 2003; Ellison *et al.* 2011). In total, we analyzed 18 genomes from 10 *Neurospora* species, of which 15 genomes were PacBio-sequenced, resulting in dense taxon sampling both across the phylogeny and within specific species (*N. sitophila* (4 strains) and *N. intermedia* (3 strains)) (Figure 1, Table S1).

The assemblies of the PacBio-sequenced genomes were nearly gapless. We recovered contigs that could be anchored to the seven chromosomes in each of the PacBio genome assemblies that were co-linear with the seven chromosomes in *N. crassa* (Perkins *et al.* 2001), indicating that our assemblies were at the chromosome level. One assembly (strain FGSC 5940) had no gaps [the one excess contig is mitochondrial DNA], and from three other strains (W1434, and FGSC 8807 and 9046) we identified up to 8 unplaced contigs that represent gaps in the assembly resulting from the repetitive ribosomal RNA genes. Two strains of *N. intermedia* (FGSC 8767 and 8761) had particularly more “gappy” assemblies, where 24 and 10 contigs

mapped to the 7 *N. crassa* chromosomes, respectively. For all PacBio assemblies, the 1 to 29 unplaced contigs that could not be mapped to the *N. crassa* reference genome were mitochondrial DNA and ribosomal RNA genes. The assembly sizes of the examined *Neurospora* genomes ranged from 37.8 to 43.9 Mb, and for *N. crassa* it was 41 Mb (Table S1).

#### *The curated Neurospora TE library*

TE identification in the nearly gapless genome assemblies involved sequence similarity searches based on a modified version of an existing query repeat sequence library (Gioti *et al.* 2012). The original 978 repeats library was curated and reduced to 844 repeats. An additional 41 *N. crassa* TE sequences obtained from NCBI GenBank, from Selker *et al.* (Selker *et al.* 2003), and from Wang *et al.* (Wang *et al.* 2015) were added to the updated TE query library for a total of 885 repeats for RepeatMasker analyses, and the library has been made available at: <https://doi.org/10.6084/m9.figshare.12199409.v1> (Nguyen and Johannesson 2021).

#### *Genomic TE content rich in LTR retrotransposons*

The *Neurospora* genomes were comprised of both Class I RNA retrotransposon and Class II DNA transposon sequences. These TE sequences represented 7.66 Mb of genomic DNA (17.43% of genome) in *N. metzenbergii*, the species with the largest genome. In two of the species with the smallest genomes, these TEs represented less than 4 Mb of the genomes (*N. sitophila*: 3.45 Mb, 9.11% of genome; and *N. discreta*: 3.95 Mb, 10.46% of genome) (Figure 1; Table S2). We found that TE sequences comprised 13.02% (5.34 Mb) of the *N. crassa* genome, which is similar to a previous report (Galagan *et al.* 2003).

The proportion of genomes occupied by the different TE families (e.g., LTR retrotransposon, LINE, or DNA elements) varied considerably among the *Neurospora* species (Figure 1, Figure S1). However, the LTR retrotransposon sequences were most abundant in all *Neurospora* genomes (Figure 1). They made up the largest fraction of the TE sequences (1.44 - 3.85 Mb or 3.74 - 8.77% of genomes), followed by LINE elements (0.34 - 1.20 Mb or 0.86 - 2.74% of genomes), DNA elements (0.09 - 0.51 Mb or 0.24 - 1.17% of genomes), and repeat sequences that could not be classified (Unclassified repeats, 0.70 - 2.06 Mb or 1.82 - 4.69% of genomes) (Table S2, Figure S1).

#### *TEs locate across all seven Neurospora chromosomes*

For all *Neurospora* genomes investigated, numerous TE sequences were identified on each of the seven *Neurospora* chromosomes, but sliding window analyses suggest certain regions per chromosome were particularly enriched ( $\log_2$  observed/expected values  $>2$ ) (Figure 2, Figure S2, top panels). Similar patterns of enrichment were observed between orthologous chromosomes in each of the *Neurospora* species (Figure S2, top panels).

#### *Genome size correlation with TE content*

In *Neurospora* species, TE content correlated positively with genome size (Figure 3), suggesting TE contribution to genome size expansion in *Neurospora*. Our data suggest that primarily LTR retrotransposons and unclassified repeats are the responsible families: while these families are correlated with larger genome sizes, other TEs were weakly or not at all correlated to genome size (Figure S3). We also see deviations from the regression, indicating that the TE expansion is not the only factor contributing to genome size variation. A number of genomes fell outside the 95% confidence interval of the regression; these genomes either had more or fewer TEs than

expected. For example, *N. discreta* and *N. sitophila* (strains FGSC 5940 and W1434, W1426) had more TE sequences than expected relative to their genome sizes, as predicted using the correlation between the remaining species (Figure 3).

#### *RIP signatures were detected in all Neurospora species*

The composite RIP index can be used as a proxy for the RIP activity in a genomic region; positive values indicate that RIP has occurred (Margolin *et al.* 1998; Lewis *et al.* 2009). For this study, we developed a method to determine genome-wide RIP indices in our assembled genomes independent of the underlying sequence context or genomic content (e.g., genic regions, repetitive regions). As expected, we observed RIP signature in the reference *N. crassa* genome, and similar patterns were observed for the other 17 genomes in this study (Figure S4, “obs”).

Composite RIP index values for each genome assembly were also calculated for 10 kb windows, delineated independently of the sequence context or genomic content, and plotted over the length of the chromosome (Figure 2, Figure S2, bottom panels). Positive values indicated that a region had experienced RIP activity. Note that these RIPed regions coincided with windows enriched in TEs (Figure 2, right panels, Figure S5). Taken together, on a genome-wide scale, TE sequences were observed to have experienced RIP.

#### *TE partly accounts for intra-species genome size variation*

We observed genome size variation between strains of *N. sitophila* and *N. intermedia* (Table S1). To better understand the variation within species, particularly whether TEs accounted for the gain or loss of DNA, we compared the high-quality nearly gapless genome assemblies of three independent *N. intermedia* strains and four independent *N. sitophila* (Table S1).

We found genome size differences among *N. intermedia* as the FGSC 8767 assembly was approximately 3.4 Mb smaller than FGSC 8807 and FGSC 8761 (Table S1), and TEs contributed to 1.95 Mb of this size difference (Table S2). The FGSC 8767 assembly was, however, more fragmented than the other two *N. intermedia* assemblies that had fewer and longer contigs mapped to *N. crassa* (Table S1). Therefore, we cannot rule out that the difference in size among the *N. intermedia* genomes is due to unassembled repetitive sequences in the FGSC 8767 genome, demonstrating the importance of high-quality and nearly gapless assemblies for more specific/detailed whole-genome comparisons within and across species.

Among the four *N. sitophila* strains, W1434 had approximately 1 Mb larger genome than the other three strains (W1426, FGSC 5940, FGSC 5941: 37.9-38.4 Mb) (Table S1). In contrast to *N. intermedia*, the genome assemblies of the *N. sitophila* strains are more even in quality, with fully mapped contigs to *N. crassa* (with the exception of mitochondrial DNA and ribosomal RNA genes). Hence, we suggest that the larger genome size of W1434 is a consequence of different genome rearrangements, for example by accumulation of certain TE families, rather than assembly quality due to fragmentation. Indeed, the size difference could be accounted for by an excess in repetitive sequences that were dispersed along the seven chromosomes (Figure S2). The TE content of W1434 was about 0.7 Mb larger than the other two *N. sitophila* genomes (4.25 Mb compared with 3.45-3.58 Mb in W1426, FGSC 5940, FGSC 5941), of which 0.589 Mb (82%) of sequences were contributed by LTR retrotransposons alone (Table S2).

#### *Lineage-specific TEs in Neurospora*

We investigated TE activity in *Neurospora* by searching for lineage-specific TE sequences. We scored pairwise alignments of individual TE-flanking regions for TE presence and absence. We

assumed a specific TE present to be the result of an insertion. These insertions indicate recent TE accumulation, and we refer to them as “pairwise lineage-specific TE insertions” herein. We found between 0 and 23 insertions per genome (Table S3) and these were either LTR and LINE retrotransposons from 82 subfamilies (Table S4). In total, we found 593 pairwise lineage-specific LTR retrotransposons in our pairwise comparisons (of which 406 putative full-length insertions; Tables S3 and S4) belonging to 76 subfamilies and 509 LINE belonging to 6 subfamilies (Tables S3 and S4) suggesting a diversification of LTRs. Most of the insertions (32%) were represented by *ncra\_Tad1\_01* LINE subfamily, followed by *ncra\_LTR\_69*, *ntet\_Tad1\_01*, *Tad1.1* (4% each), *ntet\_LTR\_18*, *ncra\_LTR\_49*, *ncra\_Tad1\_06* (3% each; Table S5).

## Discussion

Genome integrity is maintained by different mechanisms to ensure genome function and successful transmission of genetic material to the next generation. However, variation in genomic DNA content can still be observed. The study of genome content variations has been made possible by availability of high-quality genome assemblies via the employment of next generation and newer sequencing technologies, increasingly available at low costs and with increasing efficiency and accuracy. In this study, we investigated 18 nearly gapless genome assemblies of ten *Neurospora* species to determine and compare genome-wide TE distribution and their associated RIP patterns.

The first report on the 38 Mb genome assembly of *N. crassa* indicated that 10% of the genome consisted of repetitive elements, which were detected based on filtering of alignments longer than 200 bp (Galagan *et al.* 2003). In our study, we found that 13% of the genome in *N. crassa* consists of TE sequences. The difference between our studies may be due to the underrepresentation of SINE elements and truncated elements in the original draft genome. We now add to the body of knowledge that closely-related *Neurospora* species have similarly large proportions of their genomes comprised of repetitive elements. Nevertheless, we found variation in repetitive content, both between and also within species. Among our investigated *Neurospora* genomes, TE contents vary between 8.7 and 18.9% of the genome.

One main finding of this study is the correlation between TE content and genome size in *Neurospora*. Genome size has been reported to be positively correlated with the abundance of TEs in diverse lineages of eukaryotes (Kidwell 2002; Oliver *et al.* 2013). Many studies on the topic have been based on comparisons across distant and diverged lineages, but as the number of high-quality genome assemblies increase, more studies comparing closely-related species are reported (Gregory and Johnston 2008; Hu *et al.* 2011; Legrand *et al.* 2019). In fungi, a positive correlation between TE content and genome size has been reported on both long and short evolutionary time scales (Ohm *et al.* 2012; Grandaubert *et al.* 2014; Benevenuto *et al.* 2018; Badet *et al.* 2020; Haridas *et al.* 2020). In our study with *Neurospora* species that diverged from a common ancestor within the last 7 million years, we observed similar genome size-TE correlations. Our result is noteworthy given that, in the model species *N. crassa*, TEs are expected to be inactive due to the operation of RIP, which we here verify to be a conserved genome defense system in the genus. Hence, our data suggest that in spite of an active RIP system, TEs are actively contributing to genome size evolution in *Neurospora*. This pattern can be explained by TEs being able to escape RIP and proliferate in the genome, a claim that needs experimental verification. Alternatively, as previously demonstrated, the contribution of TEs to genome size evolution depends on TE accumulation relative to the overall deletion rate (Kapusta *et al.* 2017), and at this point we cannot exclude the alternative possibility that the variation in

TE content between *Neurospora* strains and species is a result of differential loss of an ancestral pool of TEs in the different lineages, possibly via genomic rearrangements (see for example (Ren *et al.* 2018)).

Large genome size variation among individuals within a species has also previously been attributed to TEs (e.g. (Kalendar *et al.* 2000) and (Black and Rai 1988)). Our study involves broad sampling of *Neurospora* species, of which two are represented by multiple strains providing a brief look into intraspecific variation also in our data. With sampling of up to four strains in *N. sitophila* and *N. intermedia*, we indeed found variation in TE content, suggesting that TEs can contribute to genome size evolution also within species. Our view of TE dynamics over short evolutionary time is supported by our identified specific TE insertions. These numbers are small, and primarily represented by LTR retrotransposons and LINE elements. We used a conservative analysis pipeline based on identification of orthologous regions flanking the TE and queried whether the corresponding region contains a TE or not (thus intact or uninterrupted by a TE) in the contrasting genome. TEs often cluster, and/or nest within other TEs over time (Jedlicka *et al.* 2019). Furthermore, RIP has been shown to extend beyond the duplicated target region (Foss *et al.* 1991; Gladyshev and Kleckner 2016), complicating identification of orthologous TE flanking sequences between the compared genomes. Thus, the low number of observed genomic pairwise TE differences could be hampered by nested insertion and RIP activities, highlighting the need for a broad sampling and sequencing to resolve most individual TE loci along the host phylogeny.

Our study demonstrates that RIP signatures can be computationally assessed in *Neurospora* similar to previous reports on a wide diversity of ascomycete and basidiomycete fungi (Selker *et al.* 2003; Lewis *et al.* 2009; Clutterbuck 2011; Horns *et al.* 2012; Clutterbuck 2017). Tools have previously been developed to study RIP in fungal genomes (Montanini *et al.* 2014; Van Wyk *et al.* 2019). Here, we developed a method to assess RIP genome-wide, independent of the underlying sequence context and genomic content, available in the R package `ripr` (<https://github.com/NBISweden/ripr>). This resource can prove useful to identify genomic regions, e.g., centromeres that are TE rich (Smith *et al.* 2012; Friedman and Freitag 2017), without identification of TEs first, which can be a difficult task depending on the methods used for detection (Lerat 2010; Goerner-Potvin and Bourque 2018). The pattern from the present study demonstrating correlation between TEs and RIP have the potential to become a useful tool in several lines of fungal genome assessments from determining TE content by analyzing a less complicated RIP index. Also, in species where repetitive sequences such as TE frequently accumulate at centromeric regions, RIP screening could facilitate identification of such TE-associated genomic regions.

Detected RIP signatures does not however directly demonstrate RIP activity, as other mutational sources present confounding factors, as observed in retroviruses where A- and T-mutations accumulate following reverse transcription (Katz and Skalka 1990; Menendez-Arias 2009). Additionally, RIP index has been developed from *N. crassa* and dinucleotide contexts in other fungal species can vary, and even include trinucleotide contexts (Ikeda *et al.* 2002; Hood *et al.* 2005; Horns *et al.* 2012), and further experiments are needed to establish RIP activity in these species. Comparison of RIP efficiency between fungal species will benefit from experiments to quantify the extent of RIP over a number of sexual generations in *Neurospora*.

## Conclusions

Genome evolution studies benefit greatly from availability of nearly gapless genome assemblies.

Here, we have generated, analyzed, and made available high-quality nearly gapless fungal (*Neurospora*) genome assemblies. We identify variation in TE content within genomes, between genomes of the same species, and among species. Hence, in spite of the conserved RIP machinery in *Neurospora*, our data suggest that TEs are likely transposing in *Neurospora* and contribute to genome size evolution, but experimental data is needed to confirm whether the TEs are active in *Neurospora*. Finally, we developed a method to determine the RIP indices across a genome assembly to identify regions rich in TE sequences without first knowing the underlying sequence context and genomic content. The rich genomic resource we herein provide enables further in-depth analyses of the features of fungal genomes.

## Methods

**Strains investigated in the study.** *Neurospora* genomes used in this study were either generated in house or collected from public databases (Table S1). All strains sequenced for this project were obtained from the Fungal Genetics Stock Center (<http://www.fgsc.net/>) and *N. sitophila* W1434, which was provided by Jacobson et al. (Jacobson *et al.* 2006). All strains used in this study (Table S1) are referred to by their FGSC identification numbers, unless otherwise noted (e.g. W1426 and W1434, which were kindly provided by D. J. Jacobson (Jacobson *et al.* 2006)). We used three publicly available, well-annotated *Neurospora* genome assemblies that included *N. crassa* (*N. crassa* OR74A version 12 (FGSC 2489) (Galagan *et al.* 2003), sequenced by Broad Institute, corrected for the assembly error detected by Galazka et al. (Galazka *et al.* 2016)), and *N. tetrasperma* (FGSC 2508 *mat A* and FGSC 2509 *mat a* (Ellison *et al.* 2011)). Other PacBio generated assemblies were previously published (Sun *et al.* 2017; Svedberg *et al.* 2018; Hosseini *et al.* 2020; Svedberg *et al.* 2021). The remaining genomes were generated following previously published PacBio sequencing and assembly protocols (Sun *et al.* 2017; Svedberg *et al.* 2018) and are reported for the first time in this study (Table S1).

**TE sequence detection.** Repetitive DNA and TEs were identified using RepeatMasker (Version 4.0.8, <http://www.repeatmasker.org>) with Dfam database of repetitive DNA families obtained 20171107 (Hubley *et al.* 2015), and RepBase (Jurka 2000), as well as a curated *Neurospora*-specific TE library generated in this study (Nguyen and Johannesson 2021).

To curate the *Neurospora*-specific TE library, we utilized an existing *Neurospora*-specific repetitive sequence library (Gioti *et al.* 2012) to identify repetitive sequences (TEs, simple repeats and low complexity sequences) in each of the *Neurospora* assemblies and collected results into and updated TE library that included TEs called from of *N. crassa* (Galagan *et al.* 2003), *N. discreta* (<http://genome.jgi-psf.org/Neudi1/>) and *N. tetrasperma* (<http://genome.jgi-psf.org/Neute1/>) and *Neurospora* species sister to the clade containing *N. discreta* (Gioti *et al.* 2012). In summary, the original TE library (Gioti *et al.* 2012) was compiled with RepeatModeler and LTRharvest, which yielded a total 978 repeats ([http://fungalignomes.org/public/neurospora/data/repeatlib/Gioti\\_neurospora\\_repeats.renamed.lib](http://fungalignomes.org/public/neurospora/data/repeatlib/Gioti_neurospora_repeats.renamed.lib)). The TE library was further curated manually because many previously LTRharvest-identified repeats were incorrect. All 978 elements were re-assessed, including 274 sequences that were previously identified by LTRharvest, and 566 sequences that were classified by RepeatModeler as “Unknowns”, which may include multicopy genes. LTRharvest-identified repeats were queried against the conserved domain database (CDD, v 3.17, 52910 PSSMs *e-value setting 0.01*.) to determine the presence of LTR retrotransposon-related protein domains (belonging to e.g. *gag*, *pol* (including RNaseH, reverse transcriptase, and

integrase), and CHROMO domain) (Marchler-Bauer *et al.* 2014; The Uniprot Consortium 2018). Sequences were kept in the curated TE library if one of these protein domains were present, with the exception of CHROMO domains which required presence of an additional *gag* or *pol* sequence. Cellular genes, with or without transposon-related domains, were removed. RepeatModeler-classified “Unknown” repeats were queried against the CDD. Repeats with sequence similarity to cellular genes were removed from the curated TE library. Sequences without similarity matches to transposon-related protein domains (or cellular genes) were kept in the library as these could represent non-autonomous transposable elements such as SINEs normally lacking identified protein domains.

RepeatModeler-derived “DNA”, “LTR”, and “LINE” repeats (n=8, 61, and 23 sequences, respectively) were confirmed for presence of transposase- [DNA] or retrotransposon- [LTR, LINE] related protein domains in CDD. RepeatModeler-derived “SINE” repeats (n=40 sequences) were confirmed in the genomic tRNA database (Data Release 18.1 (August 2019)), using tRNAscan-SE webserver (<http://trna.ucsc.edu/tRNAscan-SE/>) and SINEBase (Vassetzky and Kramerov 2012; Lowe and Chan 2016). SINE repeats with 100% sequence similarity to a tRNA gene were removed from the curated TE library. The non-redundant nucleotide database (blastn (Altschul *et al.* 1997)) and Repbase (CENSOR: (Kohany *et al.* 2006; Bao *et al.* 2015)) were also used to identify transposon similarities for a subset of repeat sequences without conclusive results.

The curated TE library used in this study can be accessed at figshare (<https://doi.org/10.6084/m9.figshare.12199409.v1>) (Nguyen and Johannesson 2021). We analyzed each *Neurospora* assembly using the updated “Gioti\_neurospora\_repeats.renamed.v2\_191004.lib library” (844 TE sequences), the available *Neurospora crassa* TEs from NCBI Genbank (9 TE sequences), as well as additional TEs described by Selker *et al.* (Selker *et al.* 2003) (28 TE sequences) and Wang *et al.* (Wang *et al.* 2015) (4 TE sequences) by using RepeatMasker (RepeatMasker version 4.0.8, run with rmbblastn version 2.6.0+, and combined database: Dfam\_Consensus 20171107 and RepBase Update 20181026). We limited our query library to include TEs from *Neurospora* and thus exclude other fungal repeats reported in RepBase (Jurka *et al.* 2005). The RepeatMasker “.tbl” outputs were used to quantify proportions of the assemblies that contained TEs. The RepeatMasker “.out” outputs were used to calculate the composite RIP index for each of the repeat sequences. Fragmented TEs were not stitched together to create intact TEs, and nested TEs were not disentangled. Here we utilized the RepeatMasker repeat class/family categorization (column 11 in the .out file: DNA, LTR, LINE, SINE, Unclassified, Simple repeat and Low complexity).

**Identification of lineage-specific TE insertions.** In order to identify lineage-specific TEs, we modified the methodology illustrated in Guichard *et al.* (Guichard *et al.* 2018) for presence and absence alignments of TE insertion flanking DNA. We considered a TE insertion to be species-specific when the flanking DNA (occupied integration site or presence state) in the first species are found unambiguously in close proximity in the second species (unoccupied pre-integration site or absence state). All the occupied integration sites for which we found multiple pre-integration sites or undetermined orthology (e.g. when the species are too evolutionary far away from one another) were classified as unresolved integration sites.

Identification involves two steps. First, we extracted 500 bp flanking the TE insertions from each *Neurospora* assembly (based on the RepeatMasker annotation) and aligned to all other *Neurospora* assemblies. The alignment was filtered for putative unoccupied pre-integration sites

by identifying TE flanking DNA that aligned less than 50 bp apart and with at least 70% similarity (to take RIP into account). Second, we required the putative pre-integration sites to be unique and therefore aligned these to the first species. If the putative pre-integration site mapped uniquely and at least for 80% of the locus length to the TE integration site, the insertion was classified as a species-specific insertion. In case the putative pre-integration sites had ambiguous alignments in the first comparison assembly, the insertion was classified as an unresolved site. Finally, when both TE flanking sequences failed to map confidently, the corresponding insertions were classified as unresolved sites. Scripts can be found in the Github repository <https://github.com/ValentinaBoP/NeurosporaSpecificTE>.

**TE distribution in the genome.** To determine the distribution of TEs along the *Neurospora* genomes, we calculated log<sub>2</sub>-ratios of observed to expected TE content in 10-kb non-overlapping sliding windows. For each window, observed TE content was calculated by summing the intersection of the window coordinates with the TE coordinates, as defined in the RepeatMasker output file “.out”. The expected TE content was calculated as the window size times the genome-wide TE fraction.

**Composite RIP index calculations.** RIP indices were calculated for each of the transposon sequences identified by Repeat Masker. Frequencies of TpA, ApT, CpA, TpG, ApC, and GpT dinucleotides in each of the sequences were tabulated in the RepeatMasker outputs “.out” file; sequences were extracted using the EMBOSS seqret function (Rice *et al.* 2000). From these frequencies, the ratios TpA/ApT (“RIP product index”) and (CpA+TpG)/(ApC+GpT) (“RIP substrate index”) and the composite RIP index [(TpA/ApT) - (CpA+TpG)/(ApC+GpT)] were calculated (Margolin *et al.* 1998; Lewis *et al.* 2009). TEs considered to have experienced RIP are described as TpA/ApT > 0.89 and (CpA+TpG)/(ApC+GpT) < 1.03, as well as positive values of the composite RIP index (Margolin *et al.* 1998; Galagan *et al.* 2003; Lewis *et al.* 2009).

To assess the significance of composite RIP scores in 10-kb non-overlapping sliding windows for TEs in a *Neurospora* genome, we first validated that the composite RIP index could be applied across a genome to indicate regions of RIP activity. We used each genome in turn as reference and generated mock genomes with equal length to the respective reference genome, and used these to calculate null distributions of RIP scores. Mock sequences were generated in two ways, either by shuffling all genome positions or drawing nucleotides at random from the observed nucleotide frequency distribution, preserving genome length in both cases. For each panel in Figure S4, three distributions were plotted. First, the observed RIP distribution for a genome was calculated using the repeats determined by RepeatMasker (“obs”). Second given the positions of the repeats in the genome, RIP scores were recalculated by shuffling the repeat positions for two different cases. A random genome was obtained by shuffling the bases (“shuffle”), and third, generated by sampling nucleotides from the observed nucleotide frequency distribution (“frequency”), Cases 2 and 3 will disrupt all repeat regions such that no RIP signal should be observed. Functions to calculate RIP were implemented in R and are available in the R package *ripr*; the analysis pipeline is also in *figshare* (<https://doi.org/10.6084/m9.figshare.12652433.v2>).

The correlation between windows with RIPed sequences and windows enriched in TEs were visualized with the *segmented* package (version 1.2-0) in R (version 3.6.0) (R Core Team 2013; Muggeo and Muggeo 2017). Linear regression models were estimated with two segmented relationships. Estimates of the slopes and breakpoints are provided. The psi (“psi1”) value

indicates the breakpoint. The adjusted  $R^2$  (“adjusted R”) value gives an indication of how well the linear models fit the data, adjusted for the number of parameters, and considers both regression lines.

### **Data and Resource Availability**

All previously unpublished raw sequencing reads generated in this study have been deposited at the Sequence Read Archive as BioProject PRJNA622402. All genome assemblies based on PacBio sequencing technology and reference assemblies, the RepeatMasker tables of repeats for all genomes (“.tbl” and “.out” files), and the curated *Neurospora* TE library have been deposited as a collection in figshare (<https://doi.org/10.6084/m9.figshare.c.4310996>) (Nguyen *et al.* 2021). The pipeline to determine RIP and TE landscape patterns, and RIP background are contained in an R script also deposited in figshare (<https://doi.org/10.6084/m9.figshare.12652433.v2>), as part of the above collection.

### **Competing Interests**

The authors declare no conflict of interest.

### **Acknowledgements**

This work was funded from a European Research Council grant under the program H2020, ERC-2014-CoG, project 648143 (SpoKiGen; HJ). DTN was partly supported by the Sven och Lilly Lawskis Foundation. PU was financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab. AS was supported by Vetenskapsrådet (2016-05139) and FORMAS (2017-01597). PJ was supported by Vetenskapsrådet (2018-03017) and FORMAS (2018-01008).

### **Authors' contributions**

DN, VP, PU analyzed data and all authors participated in data and results interpretation. DN and HJ drafted the manuscript and VP, PU, AS, PJ contributed to writing the manuscript. All authors read and approved the final manuscript.

### **Literature Cited**

- Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
- Amselem, J., M.-H. Lebrun and H. Quesneville, 2015 Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC Genomics* 16: 1-14.
- Arabidopsis Genome, I., 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.
- Badet, T., U. Oggenfuss, L. Abraham, B. A. Mcdonald and D. Croll, 2020 A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biol.* 18: 12.
- Bao, W., K. K. Kojima and O. Kohany, 2015 Replibase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6.
- Benevenuto, J., N. S. Teixeira-Silva, E. E. Kuramae, D. Croll and C. B. Monteiro-Vitorello, 2018 Comparative genomics of smut pathogens: Insights from orphans and positively

- selected genes into host specialization. *Front. Microbiol.* 9.
- Black, W. C., and K. S. Rai, 1988 Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. *Genetics Research* 51: 185-196.
- Blumenstiel, J. P., 2011 Evolutionary dynamics of transposable elements in a small RNA world. *Trends Genet.* 27: 23-31.
- Borkovich, K. A., L. A. Alex, O. Yarden, M. Freitag, G. E. Turner *et al.*, 2004 Lessons from the genome sequence of *Neurospora crassa*: Tracing the path from genomic blueprint to multicellular organism. *Microbiol. Mol. Biol. Rev.* 68: 1-108.
- Cáceres, M., J. M. A. Ranz, A. Barbadilla, M. Long and A. Ruiz, 1999 Generation of a widespread *Drosophila* inversion by a transposable element. *Science* 285: 415-418.
- Cambareri, E., M. Singer and E. Selker, 1991 Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*. *Genetics* 127: 699-710.
- Chuong, E. B., N. C. Elde and C. Feschotte, 2017 Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics* 18: 71-86.
- Clutterbuck, J. A., 2011 Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genet. Biol.* 48: 306-326.
- Clutterbuck, J. A., 2017 Genomic CG dinucleotide deficiencies associated with transposable element hypermutation in Basidiomycetes, some lower fungi, a moss and a clubmoss. *Fungal Genet. Biol.* 104: 16-28.
- Colot, V., and J.-L. Rossignol, 1999 Eukaryotic DNA methylation as an evolutionary device. *Bioessays* 21: 402-411.
- Corcoran, P., J. L. Anderson, D. J. Jacobson, Y. Sun, P. Ni *et al.*, 2016 Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus *Neurospora tetrasperma*. *Genome Res.* 26: 486-498.
- Corcoran, P., J. R. Dettman, Y. Sun, E. M. Luque, L. M. Corrochano *et al.*, 2014 A global multilocus analysis of the model fungus *Neurospora* reveals a single recent origin of a novel genetic system. *Mol. Phylogenet. Evol.* 78: 136-147.
- Corcoran, P., D. J. Jacobson, M. I. Bidartondo, P. C. Hickey, J. F. Kerekes *et al.*, 2012 Quantifying functional heterothallism in the pseudohomothallic ascomycete *Neurospora tetrasperma*. *Fungal Biol.* 116: 962-975.
- Dettman, J. R., D. J. Jacobson and J. W. Taylor, 2003 A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* 57: 2703-2720, 2718.
- Ellison, C. E., J. E. Stajich, D. J. Jacobson, D. O. Natvig, A. Lapidus *et al.*, 2011 Massive changes in genome architecture accompany the transition to self-fertility in the filamentous fungus *Neurospora tetrasperma*. *Genetics* 189: 55-69.
- Feschotte, C., and E. J. Pritham, 2007 DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41: 331-368.
- Foss, E. J., P. W. Garrett, J. A. Kinsey and E. U. Selker, 1991 Specificity of repeat-induced point mutation (RIP) in *Neurospora*: Sensitivity of non-*Neurospora* sequences, a natural diverged tandem duplication, and unique DNA adjacent to a duplicated region. *Genetics* 127: 711-717.
- Freitag, M., R. L. Williams, G. O. Kothe and E. U. Selker, 2002 A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proceedings of the National Academy of Sciences* 99: 8802-8807.

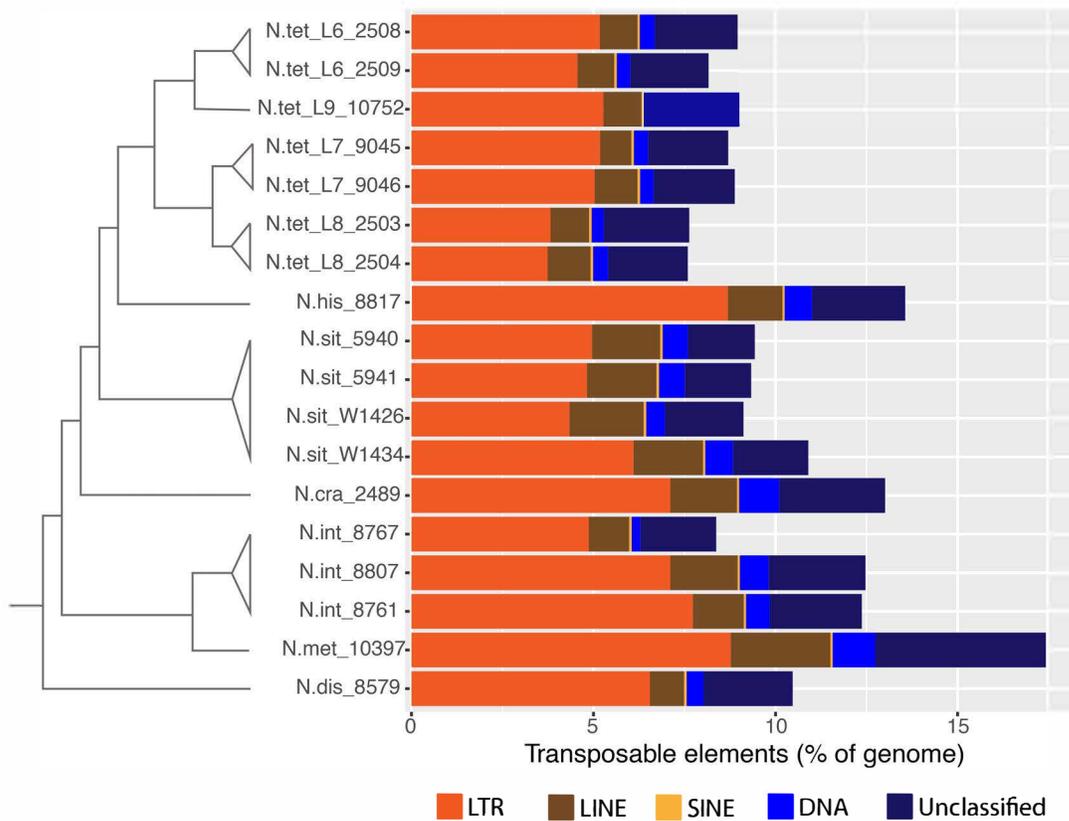
- Friedman, S., and M. Freitag, 2017 Centromeratin of fungi, pp. 85-109 in *Centromeres and Kinetochores: Discovering the Molecular Mechanisms Underlying Chromosome Inheritance*, edited by B. E. Black. Springer International Publishing, Cham.
- Galagan, J. E., S. E. Calvo, K. A. Borkovich, E. U. Selker, N. D. Read *et al.*, 2003 The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859-868.
- Galagan, J. E., and E. U. Selker, 2004 RIP: The evolutionary cost of genome defense. *Trends Genet.* 20.
- Galazka, J. M., A. D. Klocko, M. Uesaka, S. Honda, E. U. Selker *et al.*, 2016 *Neurospora* chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Res.* 26: 1069-1080.
- Gioti, A., A. A. Mushegian, R. Strandberg, J. E. Stajich and H. Johannesson, 2012 Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. *Mol. Biol. Evol.* 29: 3215-3226.
- Gladyshev, E., and N. Kleckner, 2014 Direct recognition of homology between double helices of DNA in *Neurospora crassa*. *Nature Communications* 5: 3509.
- Gladyshev, E., and N. Kleckner, 2016 Recombination-independent recognition of DNA homology for repeat-induced point mutation. *Curr. Genet.*: 1-12.
- Goerner-Potvin, P., and G. Bourque, 2018 Computational tools to unmask transposable elements. *Nature Reviews Genetics* 19: 688-704.
- Grandaubert, J., R. G. T. Lowe, J. L. Soyer, C. L. Schoch, A. P. Van De Wouw *et al.*, 2014 Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genomics* 15: 891.
- Gray, Y. H. M., 2000 It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements. *Trends Genet.* 16: 461-468.
- Gregory, T. R., and J. S. Johnston, 2008 Genome size diversity in the family Drosophilidae. *Heredity (Edinb.)* 101: 228-238.
- Grigoriev, I. V., H. Nordberg, I. Shabalov, A. Aerts, M. Cantor *et al.*, 2011 The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res.* 40: D26-D32.
- Guichard, E., V. Peona, G. Malagoli Tagliazucchi, L. Abitante, E. Jagoda *et al.*, 2018 Impact of non-LTR retrotransposons in the differentiation and evolution of anatomically modern humans. *Mobile DNA* 9: 28.
- Hane, J. K., A. H. Williams, A. P. Taranto, P. S. Solomon and R. P. Oliver, 2015 Repeat-induced point mutation: A fungal-specific, endogenous mutagenesis process, pp. 55-68 in *Genetic Transformation Systems in Fungi, Volume 2*, edited by M. A. van den Berg and K. Maruthachalam. Springer International Publishing, Cham.
- Haridas, S., R. Albert, M. Binder, J. Bloem, K. Labutti *et al.*, 2020 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. *Studies in Mycology* 96: 141-153.
- Hollister, J. D., and B. S. Gaut, 2009 Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 19: 1419-1428.
- Hood, M. E., M. Katawczik and T. Giraud, 2005 Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. *Genetics* 170: 1081-1089.
- Horns, F., E. Petit, R. Yockteng and M. E. Hood, 2012 Patterns of repeat-induced point mutation

- in transposable elements of basidiomycete fungi. *Genome Biology Evolution* 4: 240–247.
- Hosseini, S., C. Meunier, D. Nguyen, J. Reimegård and H. Johannesson, 2020 Comparative analysis of genome-wide DNA methylation in *Neurospora*. *Epigenetics*: 1-16.
- Hu, T. T., P. Pattyn, E. G. Bakker, J. Cao, J.-F. Cheng *et al.*, 2011 The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43: 476-481.
- Hubley, R., R. D. Finn, J. Clements, S. R. Eddy, T. A. Jones *et al.*, 2015 The Dfam database of repetitive DNA families. *Nucleic Acids Res.* 44: D81-D89.
- Ikeda, K., H. Nakayashiki, T. Kataoka, H. Tamba, Y. Hashimoto *et al.*, 2002 Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: Implications for its sexual cycle in the natural field context. *Mol. Microbiol.* 45: 1355-1364.
- Jacobson, D. J., J. R. Dettman, R. I. Adams, C. Boesl, S. Sultana *et al.*, 2006 New findings of *Neurospora* in Europe and comparisons of diversity in temperate climates on continental scales. *Mycologia* 98: 550-559.
- Jedlicka, P., M. Lexa, I. Vanat, R. Hobza and E. Kejnovsky, 2019 Nested plant LTR retrotransposons target specific regions of other elements, while all LTR retrotransposons often target palindromes and nucleosome-occupied regions: in silico study. *Mobile DNA* 10: 50.
- Johnson, J. M., S. Edwards, D. Shoemaker and E. E. Schadt, 2005 Dark matter in the genome: Evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* 21: 93-102.
- Johnson, L. J., 2007 The genome strikes back: The evolutionary importance of defence against mobile elements. *Evol. Biol.* 34: 121.
- Jurka, J., 2000 Repbase Update: A database and an electronic journal of repetitive elements. *Trends Genet.* 16: 418-420.
- Jurka, J., V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany *et al.*, 2005 Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110.
- Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo and A. H. Schulman, 2000 Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences* 97: 6603-6607.
- Kapusta, A., A. Suh and C. Feschotte, 2017 Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences* 114: E1460-E1469.
- Katz, R. A., and A. M. Skalka, 1990 Generation of diversity in retroviruses. *Annu. Rev. Genet.* 24: 409-445.
- Kidwell, M. G., 2002 Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115: 49-63.
- Kohany, O., A. J. Gentles, L. Hankus and J. Jurka, 2006 Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7: 474.
- Legrand, S., T. Caron, F. Maumus, S. Schwartzman, L. Quadrana *et al.*, 2019 Differential retention of transposable element-derived sequences in outcrossing *Arabidopsis* genomes. *Mobile DNA* 10: 30.
- Lerat, E., 2010 Identifying repeats and transposable elements in sequenced genomes: How to find your way through the dense forest of programs. *Heredity (Edinb.)* 104: 520-533.
- Lewis, Z. A., S. Honda, T. K. Khlafallah, J. K. Jeffress, M. Freitag *et al.*, 2009 Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. *Genome*

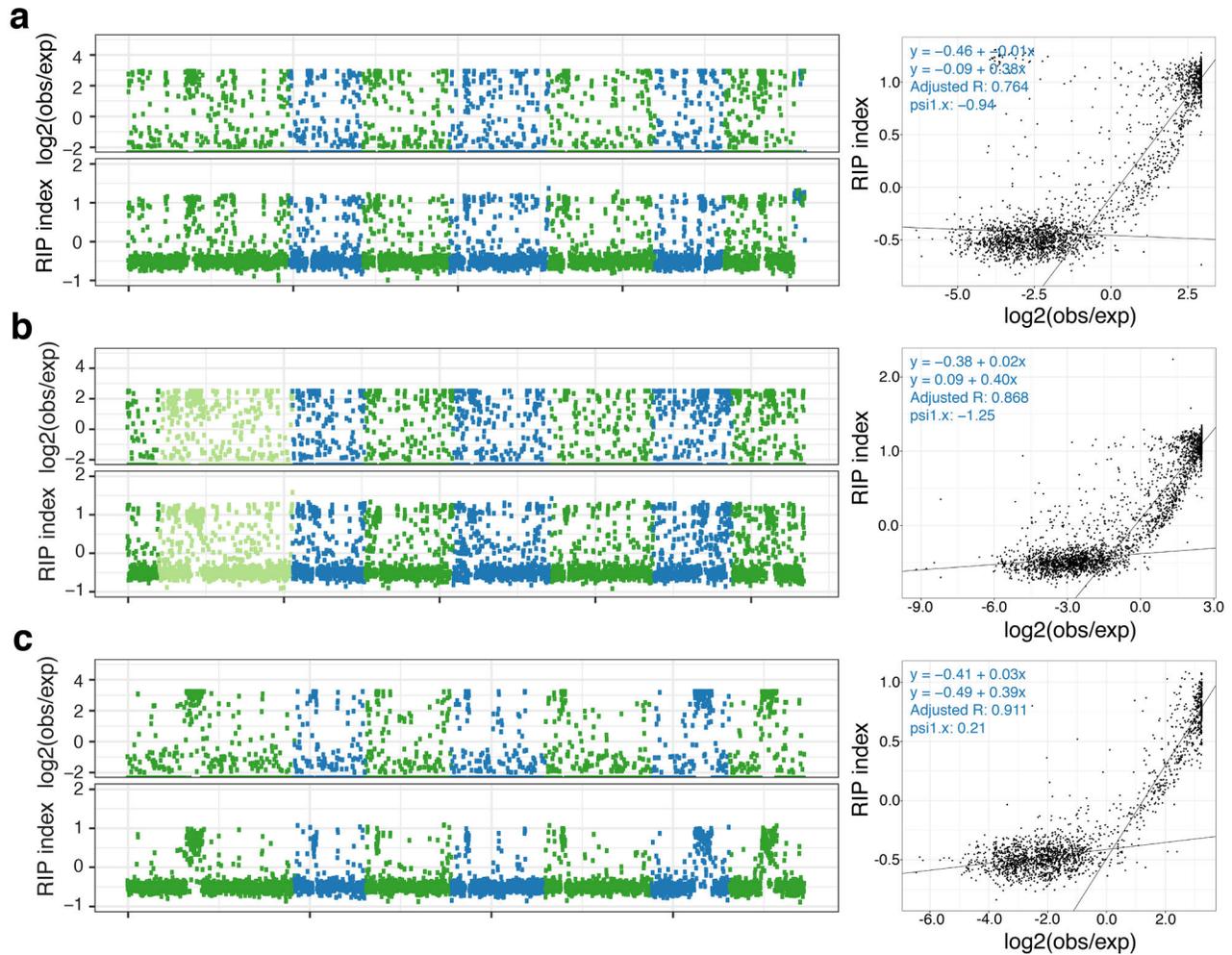
- Res. 19: 427-437.
- Lowe, T. M., and P. P. Chan, 2016 tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44: W54-W57.
- Makałowski, W., V. Gotea, A. Pande and I. Makałowska, 2019 Transposable elements: Classification, identification, and their use as a tool for comparative genomics, pp. 177-207 in *Evolutionary Genomics: Statistical and Computational Methods*, edited by M. Anisimova. Springer New York, New York, NY.
- Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz *et al.*, 2014 CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43: D222-D226.
- Margolin, B. S., P. W. Garrett-Engele, J. N. Stevens, D. Y. Fritz, C. Garrett-Engele *et al.*, 1998 A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics* 149: 1787-1797.
- Matzke, M. A., M. F. Mette and A. J. M. Matzke, 2000 Transgene silencing by the host genome defense: Implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Mol. Biol.* 43: 401-415.
- Menendez-Arias, L., 2009 Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. *Viruses* 1: 1137-1165.
- Menkis, A., E. Bastiaans, D. J. Jacobson and H. Johannesson, 2009 Phylogenetic and biological species diversity within the *Neurospora tetrasperma* complex. *J. Evol. Biol.* 22: 1923-1936.
- Montanini, B., P.-Y. Chen, M. Morselli, A. Jaroszewicz, D. Lopez *et al.*, 2014 Non-exhaustive DNA methylation-mediated transposon silencing in the black truffle genome, a complex fungal genome with massive repeat element content. *Genome Biol.* 15: 411.
- Muggeo, V. M., and M. V. M. Muggeo, 2017 Package 'segmented'. *Biometrika* 58: 516.
- Nguyen, D., and H. Johannesson, 2021 Transposable element sequence library for *Neurospora* species, figshare <https://doi.org/10.6084/m9.figshare.12199409.v1>.
- Nguyen, D., V. Peona, P. Unneberg, A. Suh, P. Jern *et al.*, 2021 Analysis pipelines for Nguyen *et al.*, (2022) "Transposons and genome dynamics in the fungal genus *Neurospora*: insights from nearly gapless genome assemblies", accepted for publication in *Fungal Genetics Reports*, figshare <https://doi.org/10.6084/m9.figshare.c.4310996>.
- Nygren, K., R. Strandberg, A. Wallberg, B. Nabholz, T. Gustafsson *et al.*, 2011 A comprehensive phylogeny of *Neurospora* reveals a link between reproductive mode and molecular evolution in fungi. *Mol. Phylogenet. Evol.* 59: 649-663.
- O'leary, N. A., M. W. Wright, J. R. Brister, S. Ciufu, D. Haddad *et al.*, 2015 Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44: D733-D745.
- Ohm, R. A., N. Feau, B. Henrissat, C. L. Schoch, B. A. Horwitz *et al.*, 2012 Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. *PLoS Pathog.* 8: e1003037.
- Oliver, K. R., J. A. Mccomb and W. K. Greene, 2013 Transposable elements: Powerful contributors to angiosperm evolution and diversity. *Genome Biol. Evol.* 5: 1886-1901.
- Peona, V., M. H. Weissensteiner and A. Suh, 2018 How complete are "complete" genome assemblies?—An avian perspective. *Mol. Ecol. Resour.* 18: 1188-1195.
- Perkins, D. D., A. Radford and M. S. Sachs, 2001 *The Neurospora Compendium: Chromosomal Loci*. Academic Press, San Diego, California.
- R Core Team, 2013 R: A language and environment for statistical computing. R Foundation for

- Statistical Computing, Vienna, Austria.
- Ren, L., W. Huang, E. K. S. Cannon, D. J. Bertioli and S. B. Cannon, 2018 A Mechanism for Genome Size Reduction Following Genomic Rearrangements. *Frontiers in Genetics* 9.
- Rice, P., I. Longden and A. Bleasby, 2000 EMBOSS: The European molecular biology open software suite. *Trends Genet.* 16: 276-277.
- Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei *et al.*, 2009 The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326: 1112.
- Selker, E. U., 1990 Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* 24: 579-613.
- Selker, E. U., 2002 Repeat-induced gene silencing in fungi, pp. 439-450 in *Advances in Genetics*, edited by C. D. Jay and C.-t. Wu. Academic Press.
- Selker, E. U., 2004 Genome defense and DNA methylation in *Neurospora*. *Cold Spring Harb. Symp. Quant. Biol.* 69: 119-124.
- Selker, E. U., N. A. Tountas, S. H. Cross, B. S. Margolin, J. G. Murphy *et al.*, 2003 The methylated component of the *Neurospora crassa* genome. *Nature* 422: 893-897.
- Singer, M. J., E. A. Kuzminova, A. Tharp, B. S. Margolin and E. U. Selker, 1995 Different frequencies of RIP among early vs. late ascospores of *Neurospora crassa*. *Fungal Genetics Reports* 42: 74-75.
- Smith, K. M., J. M. Galazka, P. A. Phatale, L. R. Connolly and M. Freitag, 2012 Centromeres of filamentous fungi. *Chromosome Res.* 20: 635-656.
- Sun, Y., J. Svedberg, M. Hiltunen, P. Corcoran and H. Johannesson, 2017 Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nature Communications* 8: 1140.
- Svedberg, J., S. Hosseini, J. Chen, A. A. Vogan, I. Mozgova *et al.*, 2018 Convergent evolution of complex genomic rearrangements in two fungal meiotic drive elements. *Nature Communications* 9: 4242.
- Svedberg, J., A. A. Vogan, N. A. Rhoades, D. Sarmarajeeva, D. J. Jacobson *et al.*, 2021 An introgressed gene causes meiotic drive in *Neurospora sitophila*. *Proceedings of the National Academy of Sciences* 118.
- The Uniprot Consortium, 2018 UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 47: D506-D515.
- Van Wyk, S., C. H. Harrison, B. D. Wingfield, L. De Vos, N. A. Van Der Merwe *et al.*, 2019 The RIPper, a web-based tool for genome-wide quantification of repeat-induced point (RIP) mutations. *PeerJ* 7: e7447.
- Vassetzky, N. S., and D. A. Kramerov, 2012 SINEBase: A database and tool for SINE analysis. *Nucleic Acids Res.* 41: D83-D89.
- Villalta, C. F., D. J. Jacobson and J. W. Taylor, 2009 Three new phylogenetic and biological *Neurospora* species: *N. hispaniola*, *N. metzenbergii* and *N. perkinsii*. *Mycologia* 101: 777-789.
- Wang, Y., K. M. Smith, J. W. Taylor, M. Freitag and J. E. Stajich, 2015 Endogenous small RNA mediates meiotic silencing of a novel DNA transposon. *G3: Genes|Genomes|Genetics* 5: 1949-1960.
- Watters, M. K., T. A. Randall, B. S. Margolin, E. U. Selker and D. R. Stadler, 1999 Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. *Genetics* 153: 705-714.
- Werren, J. H., 2011 Selfish genetic elements, genetic conflict, and evolutionary innovation.

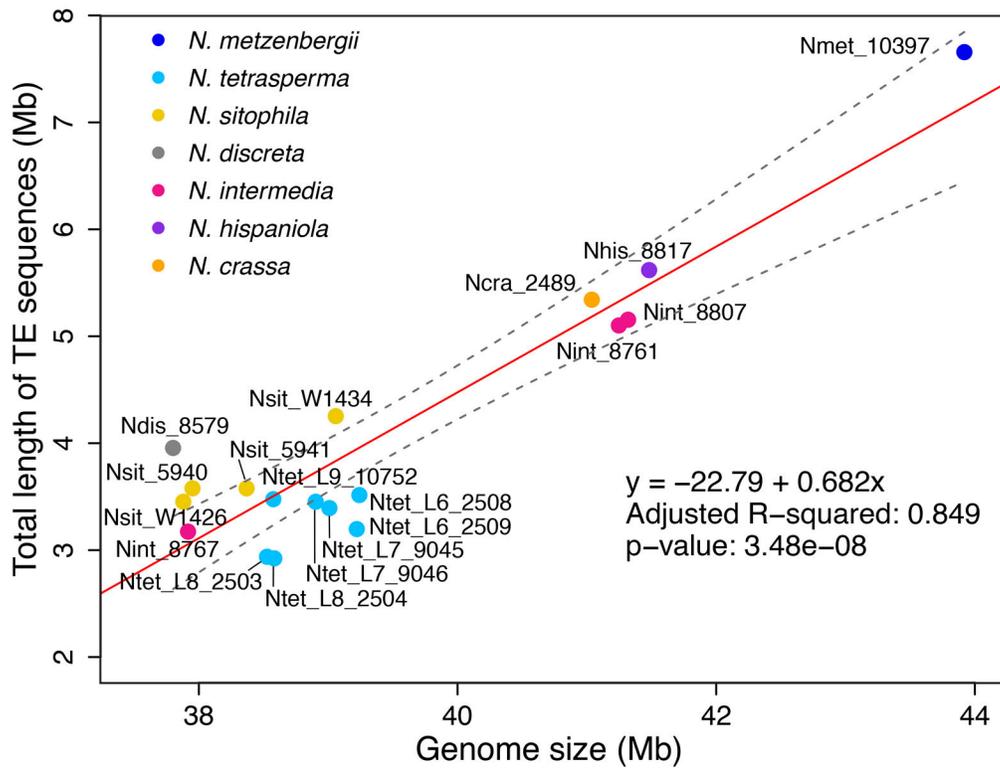
- Proceedings of the National Academy of Sciences 108: 10863-10870.
- Whittle, C. A., and H. Johannesson, 2012 *Neurospora* as a model to empirically test central hypotheses in eukaryotic genome evolution. *Bioessays* 34: 934-937.
- Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy *et al.*, 2007 A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8: 973-982.
- Zeng, J., V. K. Gupta, Y. M. Jiang, B. Yang, L. Gong *et al.*, 2019 Cross-kingdom small RNAs among animals, plants and microbes. *Cells* 8.



**Figure 1.** Schematic *Neurospora* phylogeny and TE-distributions. The left panel phylogeny was modified from (Nygren *et al.* 2011; Corcoran *et al.* 2014; Corcoran *et al.* 2016). The right panel indicates TE composition in each genome determined by calculating the number of TE nucleotides relative to the assembly length (in percent) for each of the TE families (Class I: LTR, SINE, and LINE; Class II: DNA; Unclassified).



**Figure 2.** Genome-wide TE landscape and RIP signature along *Neurospora* chromosomes. a) *Neurospora crassa* (model organism), b) *N. metzenbergii* (strain with largest genome), c) *N. discreta* (strain with one of the smallest genomes). Top panel for each species indicates the enrichment of transposable element (TE) sequences ( $\log_2(\text{observed repeat (bp)}/\text{expected repeats (bp)})$ ) determined in 10-kb windows. Each dot represents a 10-kb window. Values above 2 are herein described as enriched in TE sequences. Bottom panel for each species indicates the genome-wide composite RIP index, independent of underlying genomic content, determined in 10-kb windows. Each dot represents a 10-kb window. Positive values were herein described as a window contained sequences that experienced RIP mutation. For all plots, the alternating colors between blue and green indicate the alternation between chromosomes, with chromosome numbers following the alignment to *N. crassa*. The lighter shading indicates the presence of multiple contigs for the respective chromosome. Right panel for each species indicates the correlation between a 10-kb window for TEs and the corresponding 10-kb window for composite RIP index. Regression lines were fit using a segmented linear regression model in which the breakpoint (“psi”) is also estimated. The top regression line is the estimated fit for  $\log_2$  scores  $<$  psi value (left line) and the bottom regression line is the estimated fit for  $\log_2$  scores  $>$  psi value (right line). The adjusted  $R^2$  (“adjusted R”) value calculates fit of the regression lines to the data.



**Figure 3.** *Neurospora* genome size correlation with TE content. Total genomic TE lengths (in base pairs) correlated positively with total *Neurospora* genome assembly lengths. Each *Neurospora* species was differentially color coded (except *N. tetrasperma* for which all lineages/species are shown in light blue).

## Supplementary Information

# Transposon- and genome dynamics in the fungal genus *Neurospora*: insights from nearly gapless genome assemblies

### **Authors:**

Diem Nguyen<sup>1</sup>, Valentina Peona<sup>1</sup>, Per Unneberg<sup>2</sup>, Alexander Suh<sup>1,3</sup>, Patric Jern<sup>4</sup>, Hanna Johannesson<sup>1</sup>

### **Affiliations:**

<sup>1</sup> Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, SE-75236 Uppsala, Sweden

<sup>2</sup> Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, SE-75237 Uppsala, Sweden

<sup>3</sup> School of Biological Sciences, University of East Anglia, Norwich Research Park, NR4 7TU, Norwich, United Kingdom

<sup>4</sup> Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, SE-75123 Uppsala, Sweden

### Supplementary Figure Legends

**Figure S1.** Distribution of transposable element (TE) sequences in *Neurospora* genomes. a) Composition of TEs in each genome was determined by calculating the number of TE nucleotides relative to the genome length (reflected as percent) for each of the TE families: Class I RNA retroelements which were divided into LTR, SINE, LINE; Class II DNA transposons; and TEs that were unclassified (“Unknown”). b) Total length of TEs in each genome (bp) was plotted for each of the TE families.

**Figure S2.** TE landscape distribution in *Neurospora* genomes and genome-wide RIP pattern in *Neurospora* genomes. Top panels. Enrichment of transposable element (TE) sequences ( $\log_2(\text{observed repeat (bp)}/\text{expected repeats (bp)})$ ) were plotted for each *Neurospora* genome in 10 kb windows. Each dot represented a 10-kb window. Values above 2 were herein described as a window enriched in TE sequences. Bottom panels. Genome-wide composite RIP index, irrespective of underlying genomic content, were determined in 10 kb windows, using a custom script. Each dot represented a 10-kb window. Positive values were herein described as a window contained sequences that experienced RIP. For these plots, the alternating colors between blue and green indicate the alternation between chromosomes, following the alignment to *N. crassa*. The lighter shading indicated the presence of multiple contigs for the respective chromosome.

**Figure S3.** Correlations between total length of TE families and other repeat sequences, and *Neurospora* genome size. Total lengths of a) DNA, b) LTR, c) LINE, d) SINE, e) Unknown elements, f) low complexity, and g) simple repeat sequences were plotted against the total length of *Neurospora* genomes. Each species was color coded, and for different lineages of

## Supplementary Information

*N. tetrasperma*, they were grouped together. Each bubble was associated with a number, relating strain information: **1**, *N. tetrasperma* L6 FGSC 2508; **2**, *N. tetrasperma* L6 FGSC 2509; **3**, *N. tetrasperma* L9 FGSC 10752; **4**, *N. tetrasperma* L7 FGSC 9045; **5**, *N. tetrasperma* L7 FGSC 9046; **6**, *N. tetrasperma* L8 FGSC 2503; **7**, *N. tetrasperma* L8 FGSC 2504; **8**, *N. sitophila* FGSC 5940; **9**, *N. sitophila* FGSC 5941; **10**, *N. sitophila* W1426; **11**, *N. sitophila* W1434; **12**, *N. hispaniola* FGSC 8817; **13**, *N. crassa* FGSC 2489; **14**, *N. intermedia* FGSC 8767; **15**, *N. intermedia* FGSC 8807; **16**, *N. intermedia* FGSC 8761; **17**, *N. metzenbergii* FGSC 10397; **18**, *N. discreta* FGSC 8579.

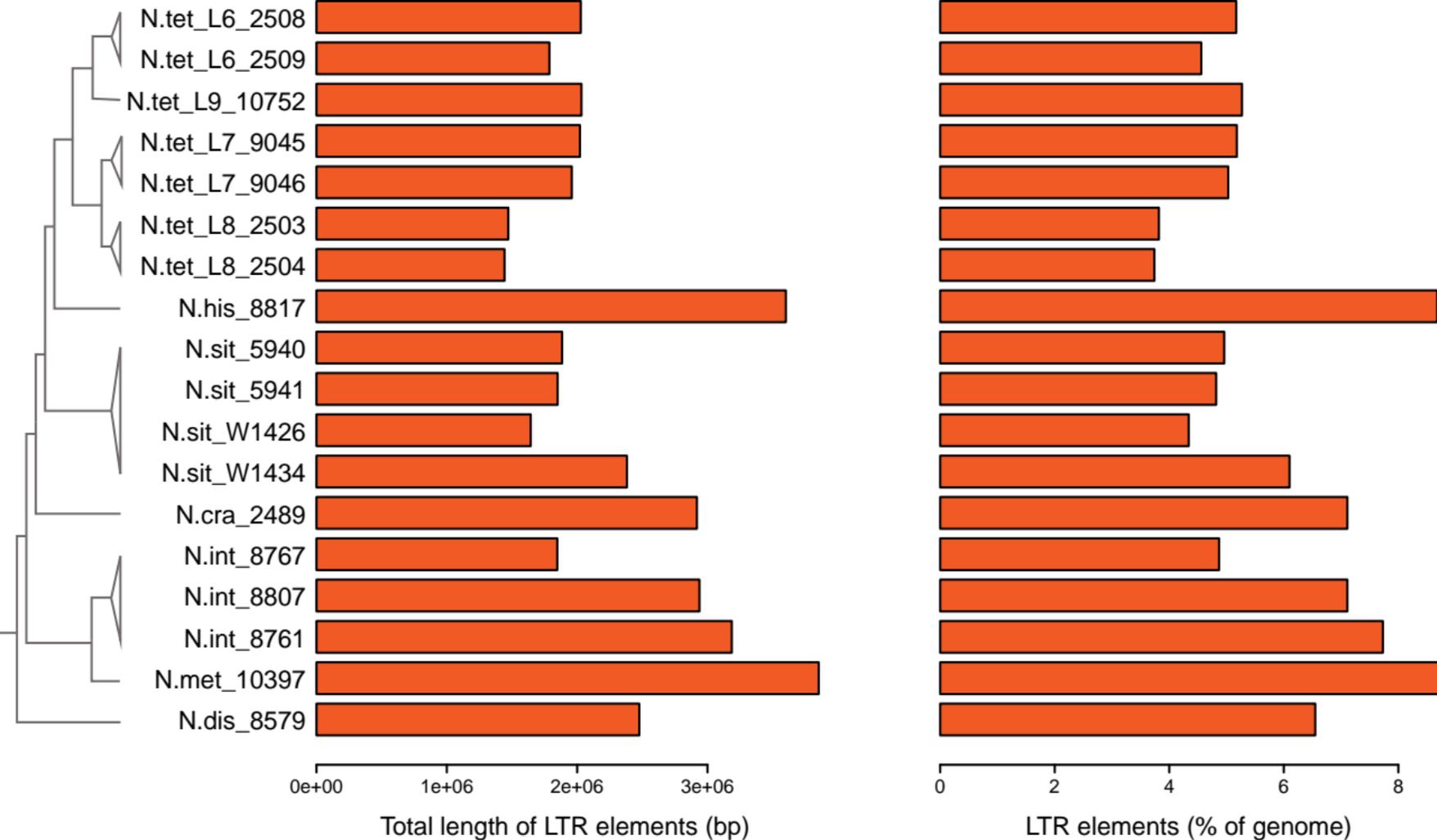
**Figure S4.** Background level of RIP signal was different from RIP signal detected at TE sequences. RIP scores were calculated for a *Neurospora* genome compared to mock sequences, i.e. mock sequences representing a background. For each panel, three distributions are plotted: 1) the observed RIP distribution for a genome calculated on the repeats determined by RepeatMasker (“obs”, purple distribution). Then, RIP scores were recalculated by randomly placing the observed repeat intervals onto mock genomes that were generated in two different ways: 2) a random genome obtained by shuffling the bases (“shuffle”, blue distribution) and 3) generated by sampling nucleotides from the observed nucleotide frequency distribution (“frequency”, yellow distribution). Cases 2 and 3 will disrupt all repeat regions such that no RIP signal should be observed.

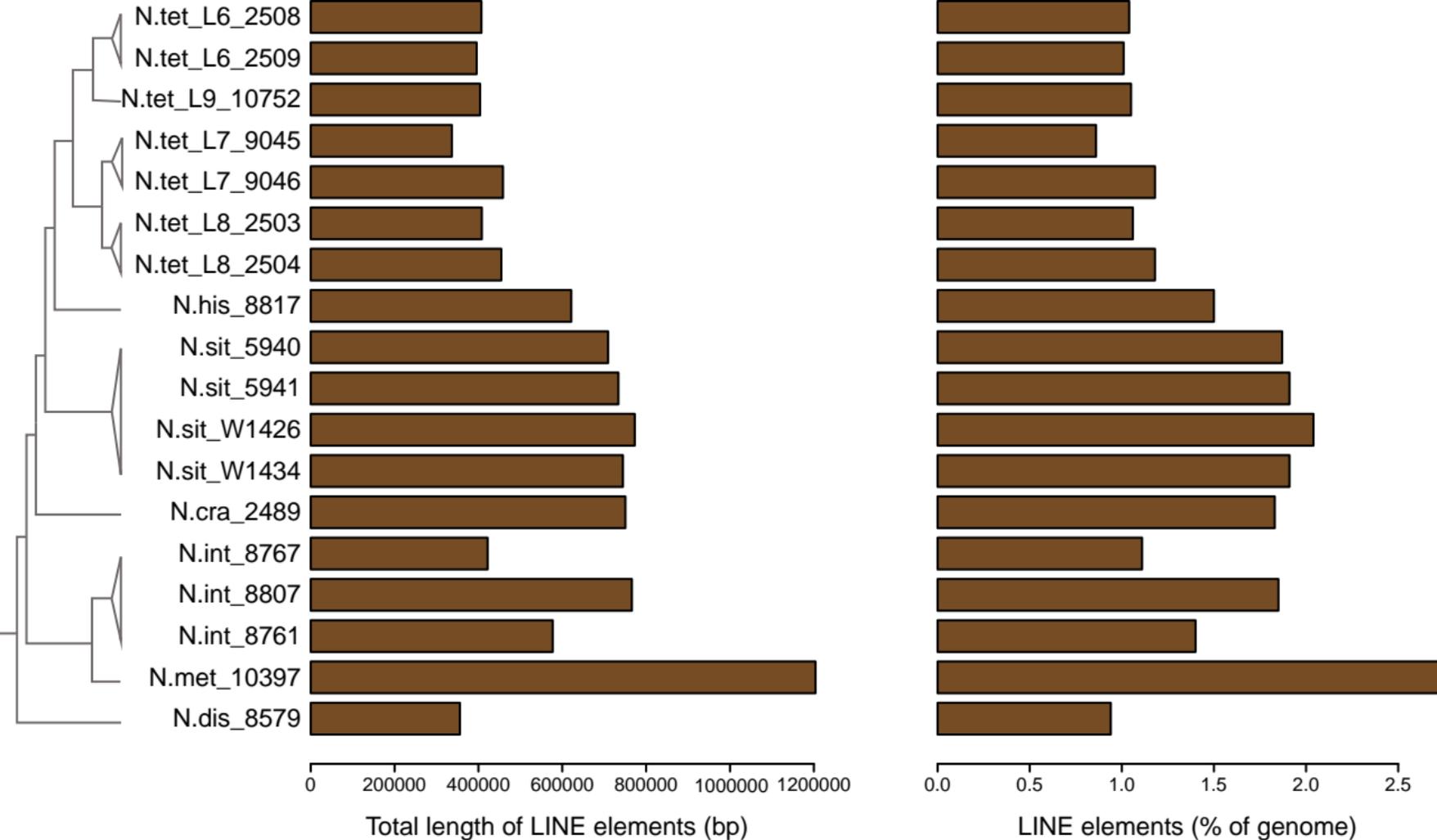
**Figure S5.** Linear correlation between enriched TE windows and RIP-positive windows in *Neurospora* genomes. For each of the 18 *Neurospora* genomes, enrichment of transposable element (TE) sequences ( $\log_2(\text{observed repeat (bp)}/\text{expected repeats (bp)})$ ) were plotted for each *Neurospora* genome in 10 kb windows. Positive values indicated that the window was enriched in TEs. Genome-wide composite RIP index, irrespective of underlying genomic

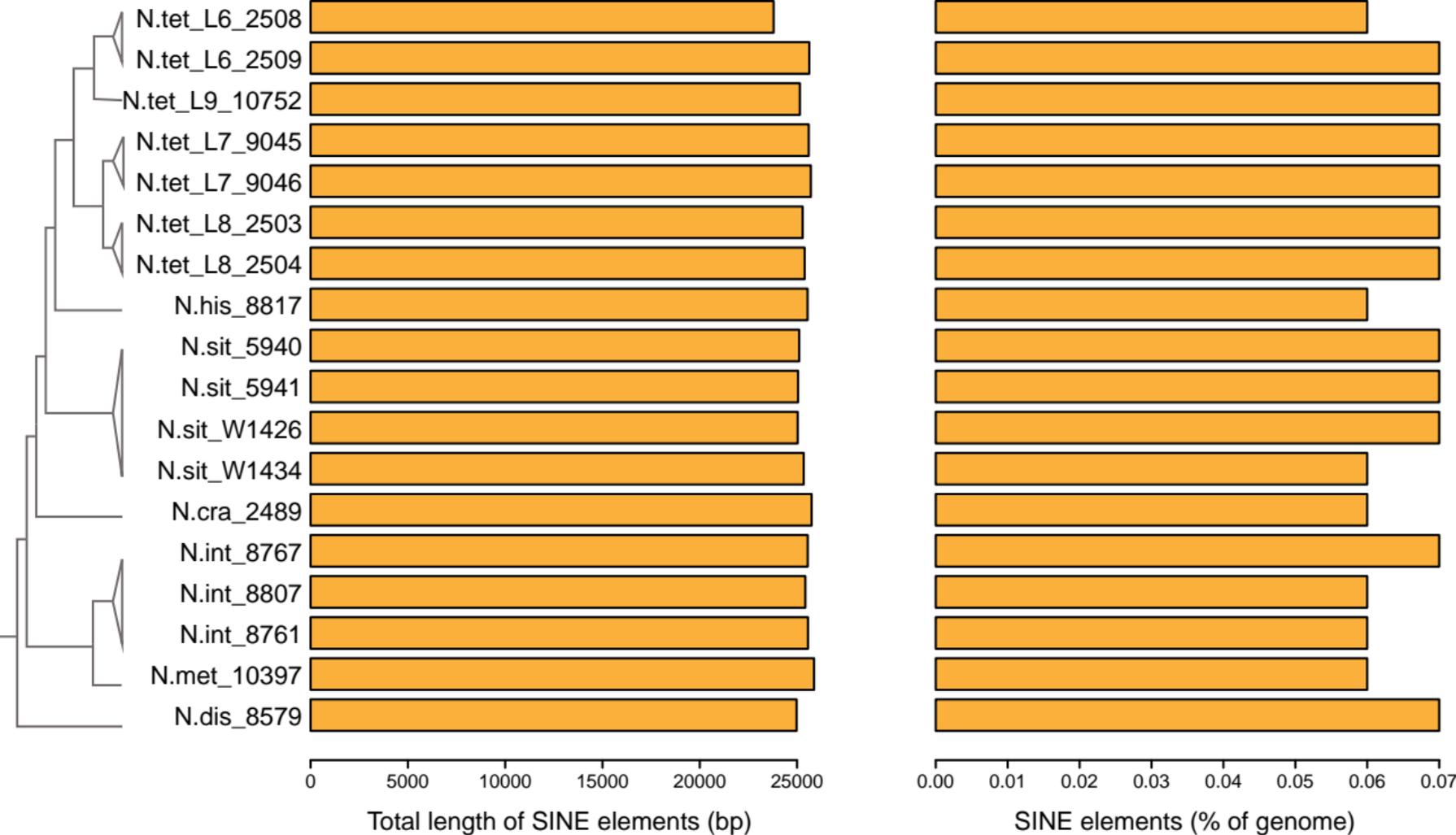
## Supplementary Information

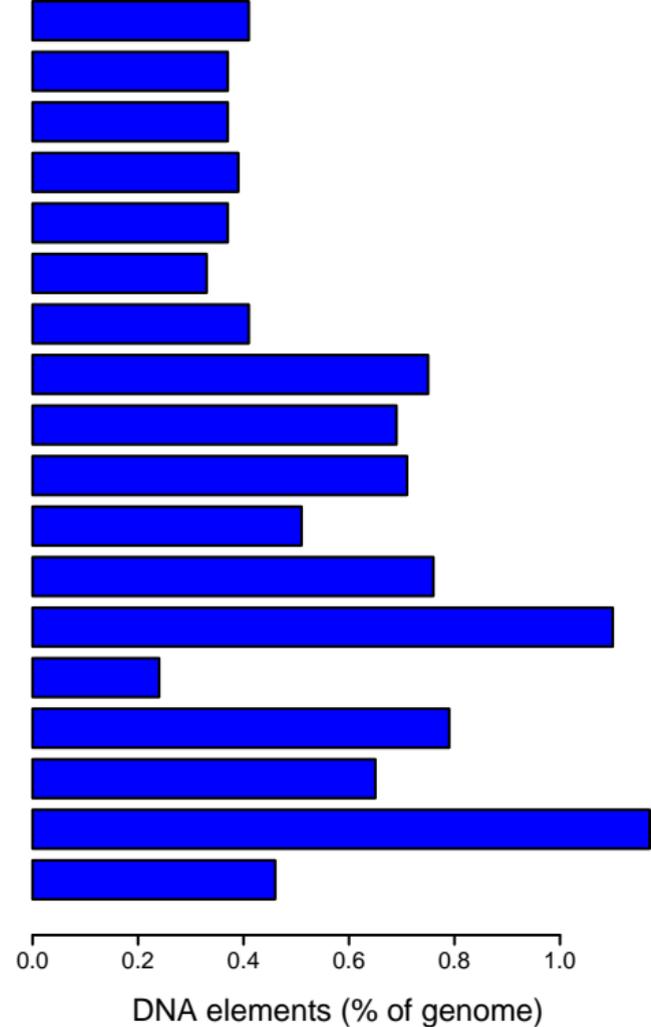
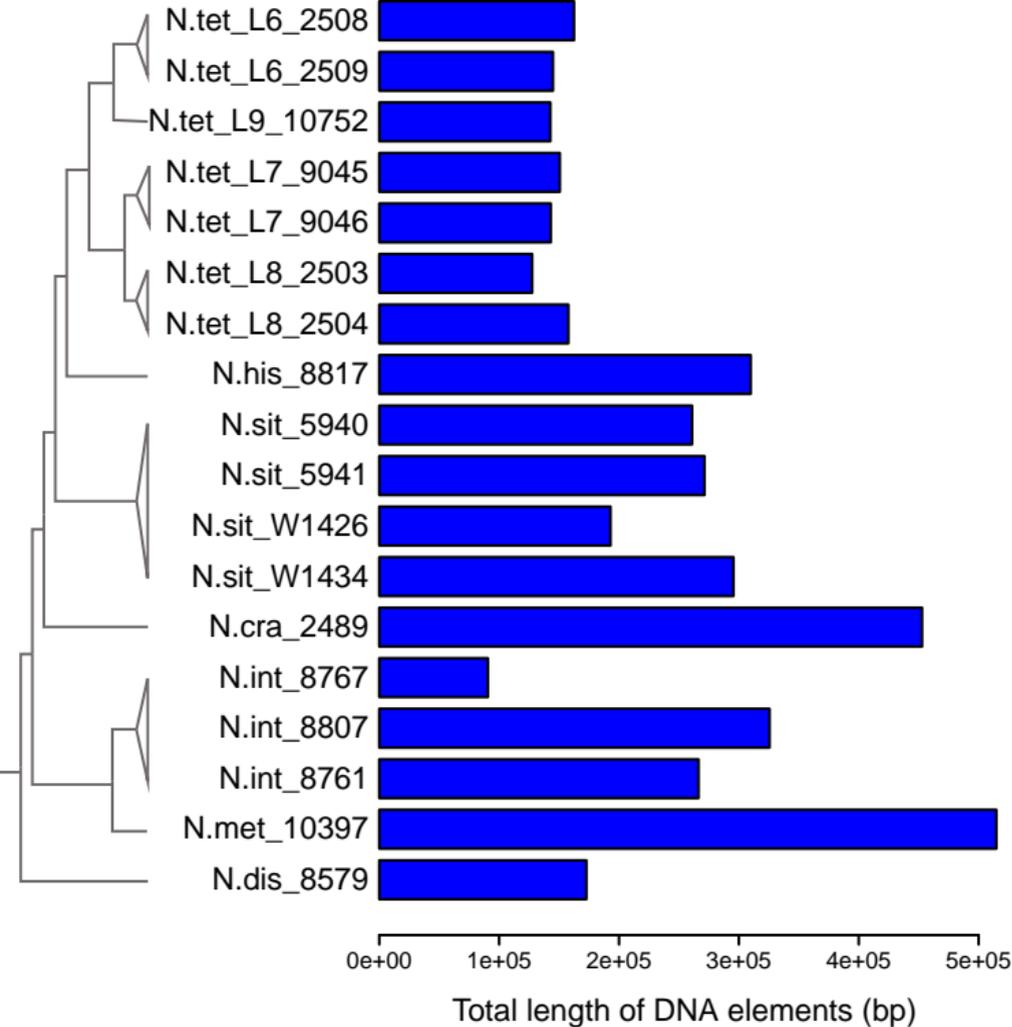
content, were determined in 10 kb windows, using a custom script. Positive values were herein described as a window contained sequences that experienced RIP. The regression lines were separately calculated for  $\log_2$  scores  $<0$  (top regression, left line) and  $\log_2$  scores  $>0$  (bottom regression, right line), with the adjusted R value calculated for the bottom regression, right line when the  $\log_2$  scores  $>0$ .

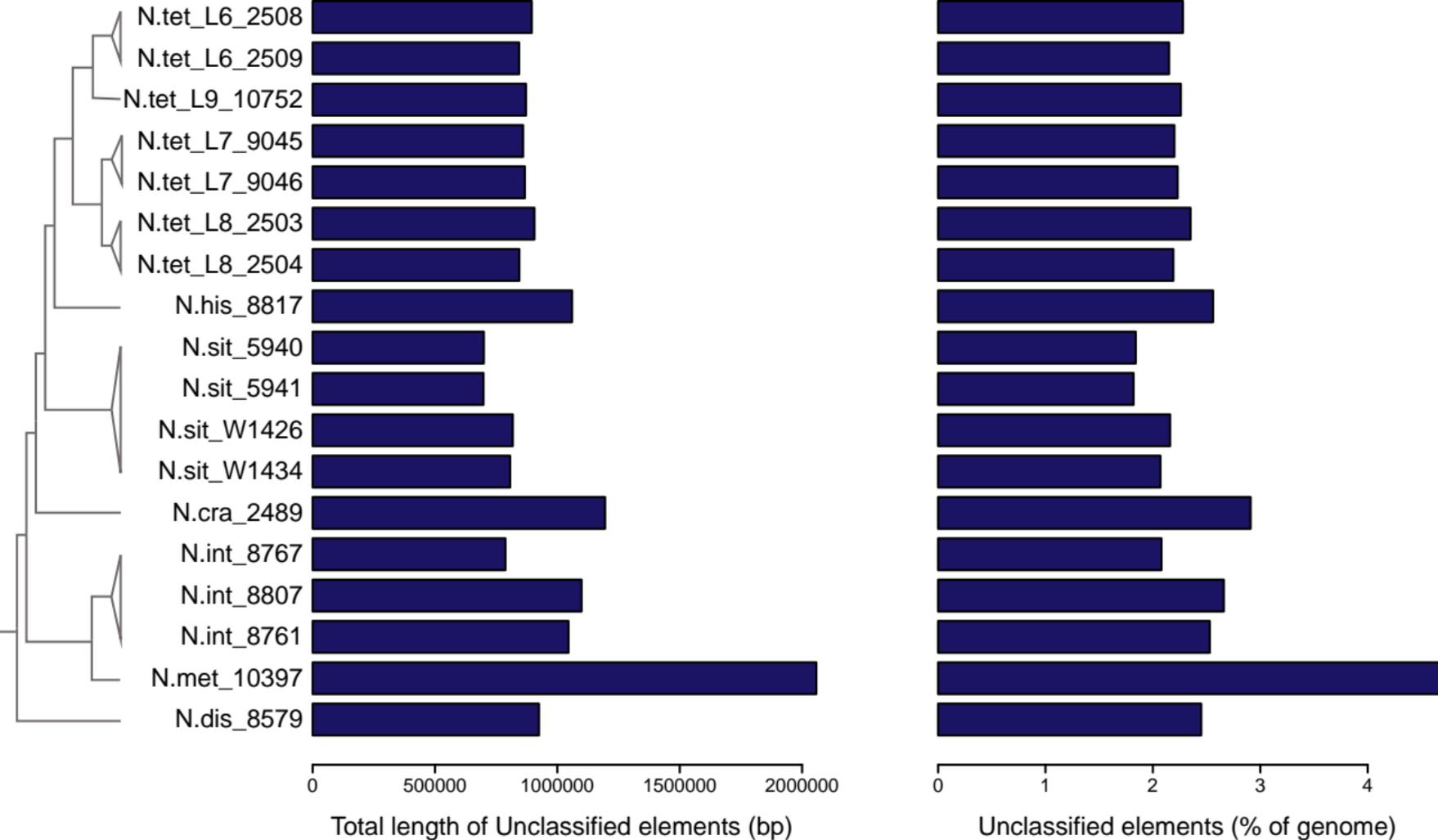
Figure S1

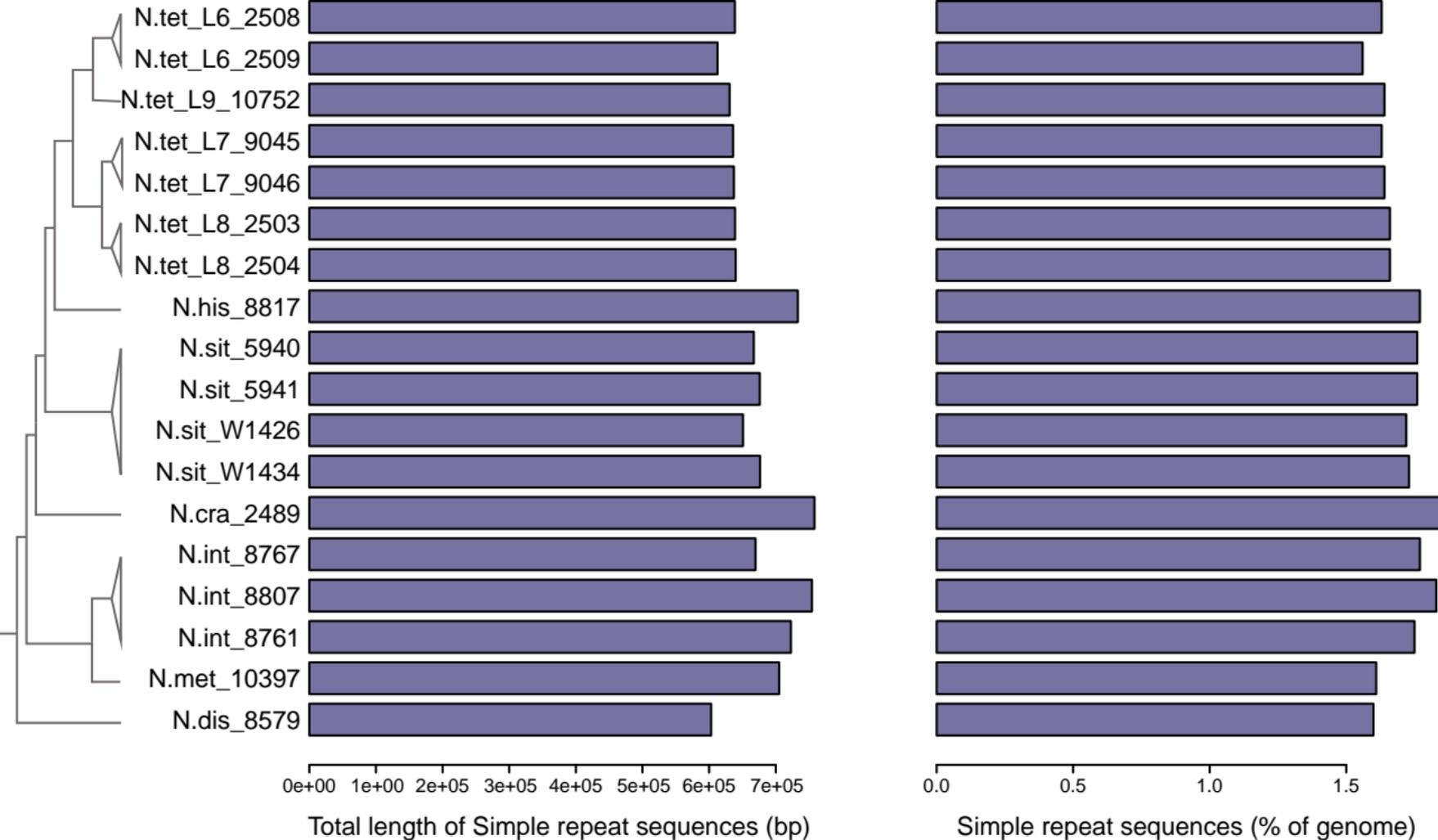












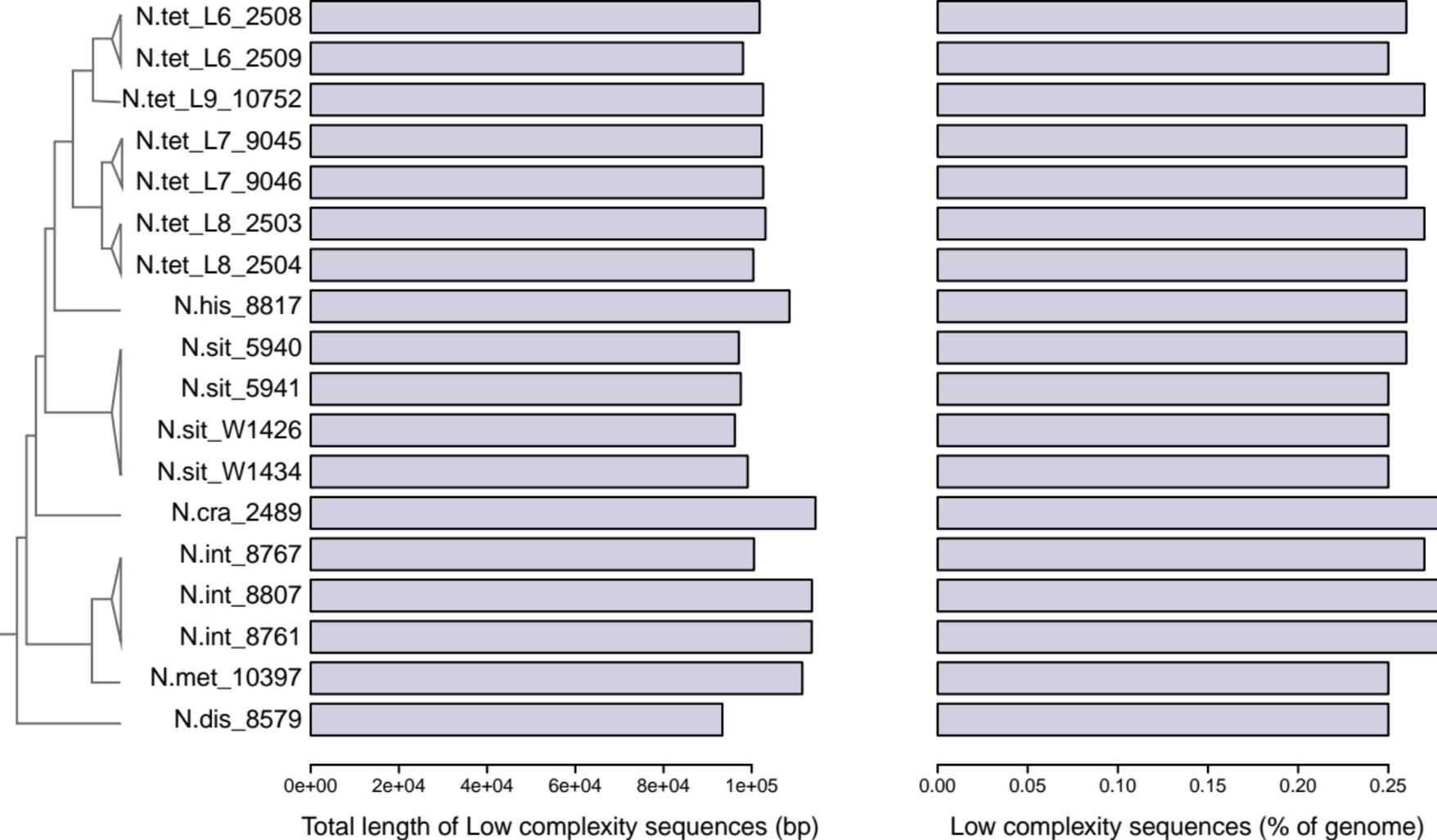
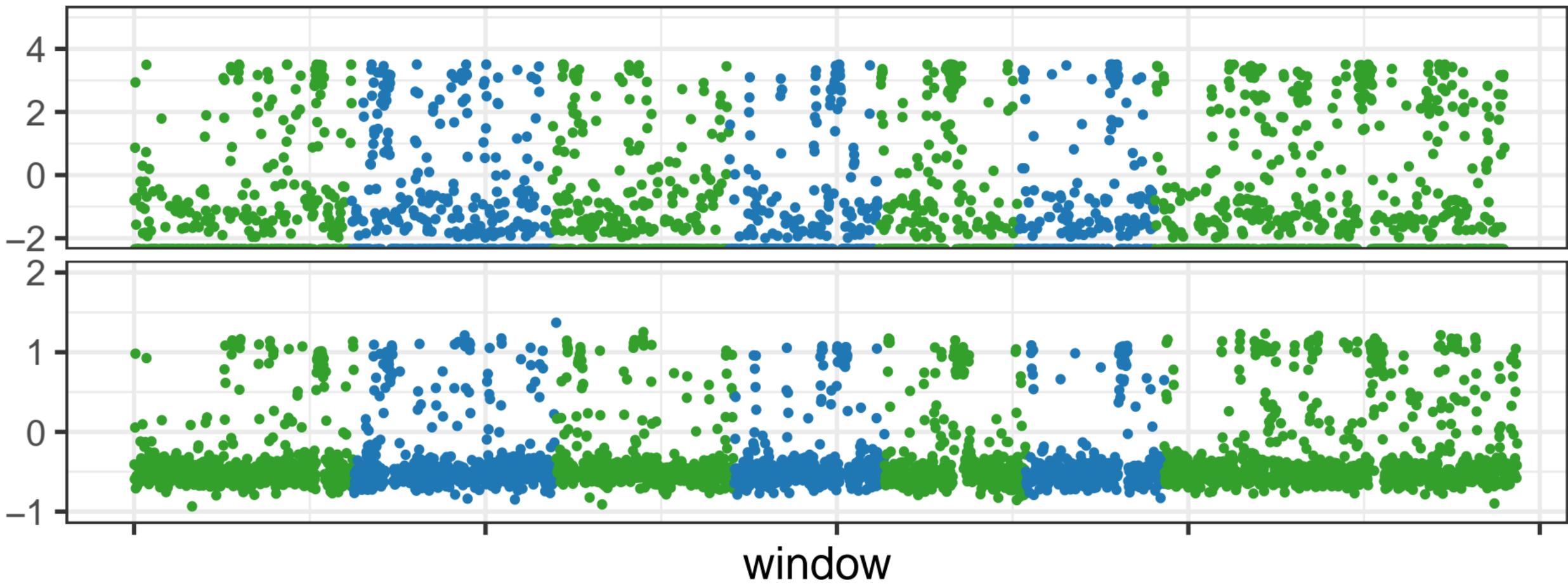


Figure S2

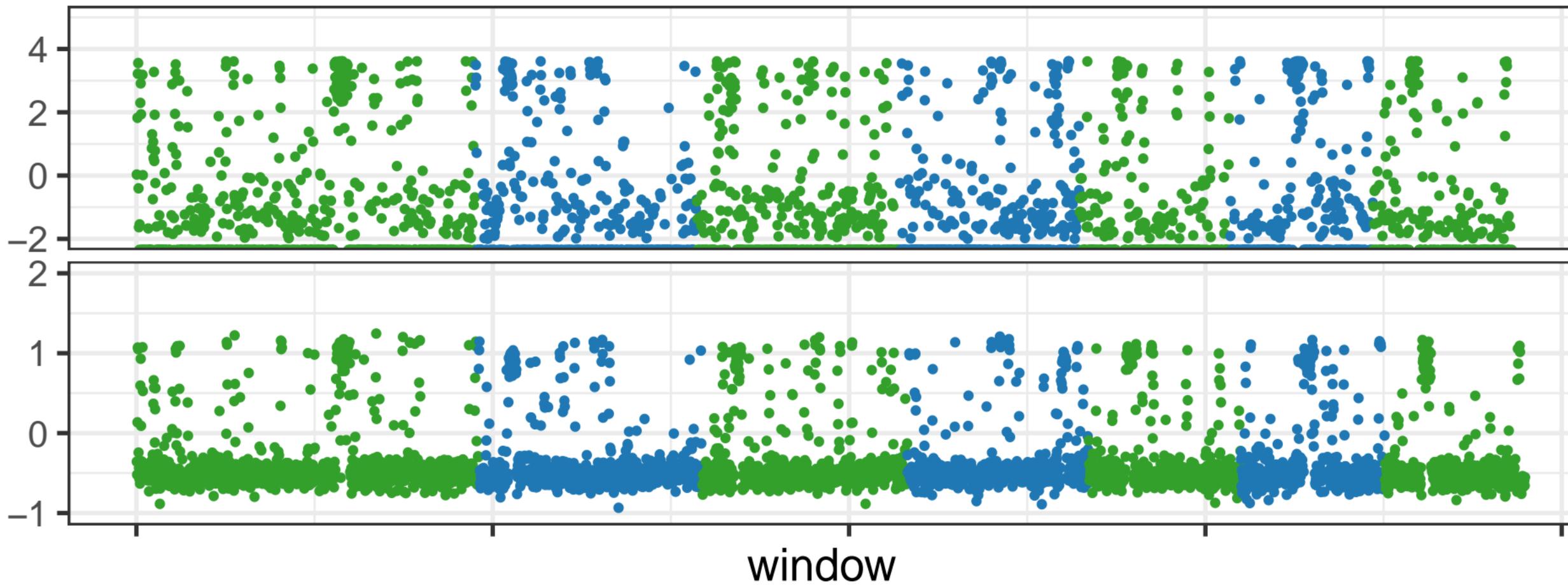
N.tet\_L6\_2508

RIP index  $\log_2(\text{obs}/\text{exp})$



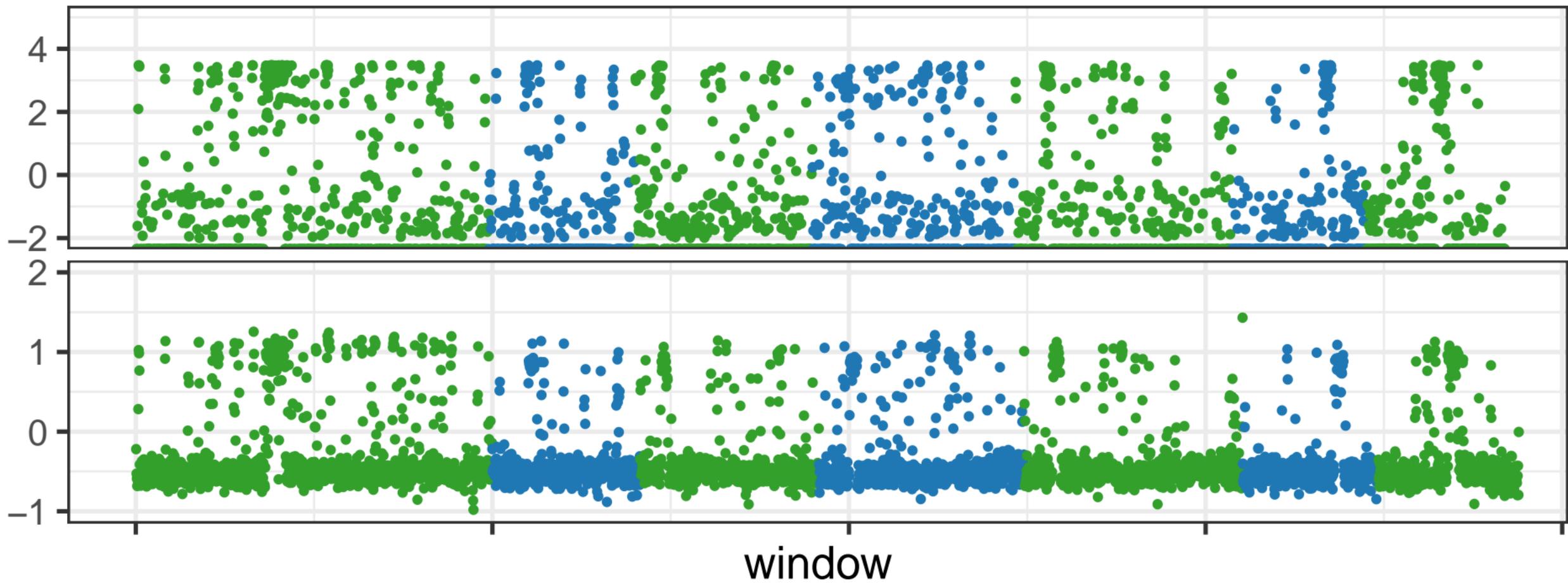
N.tet\_L6\_2509

RIP index  $\log_2(\text{obs}/\text{exp})$



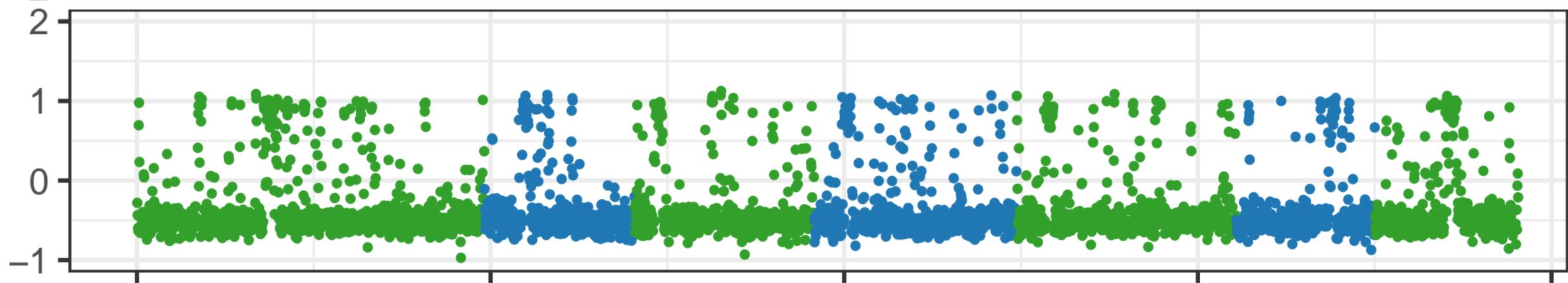
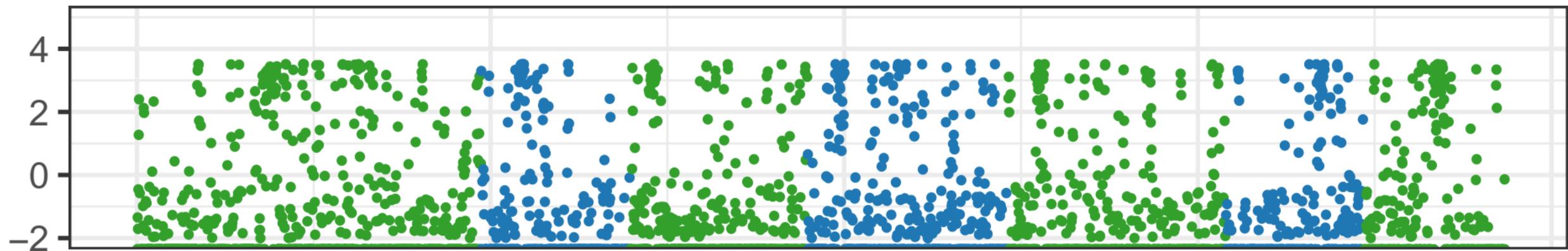
N.tet\_L9\_10752

RIP index  $\log_2(\text{obs}/\text{exp})$



N.tet\_L7\_9045

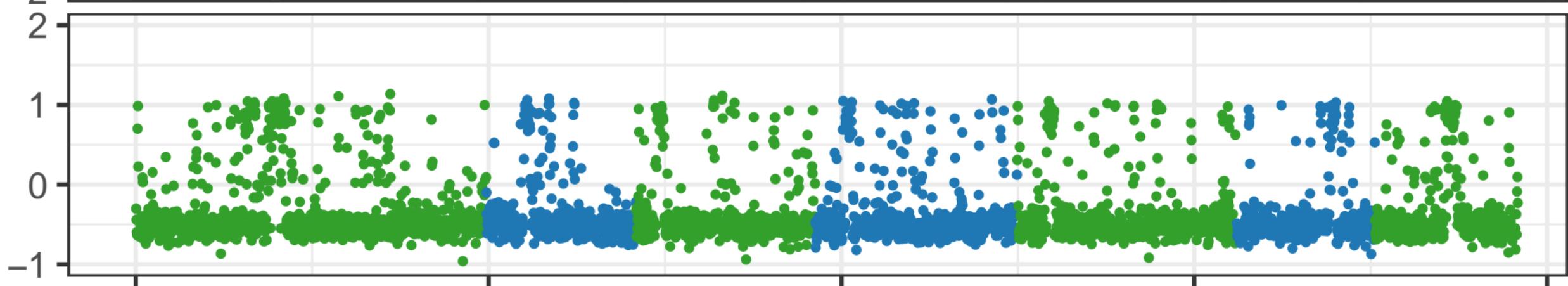
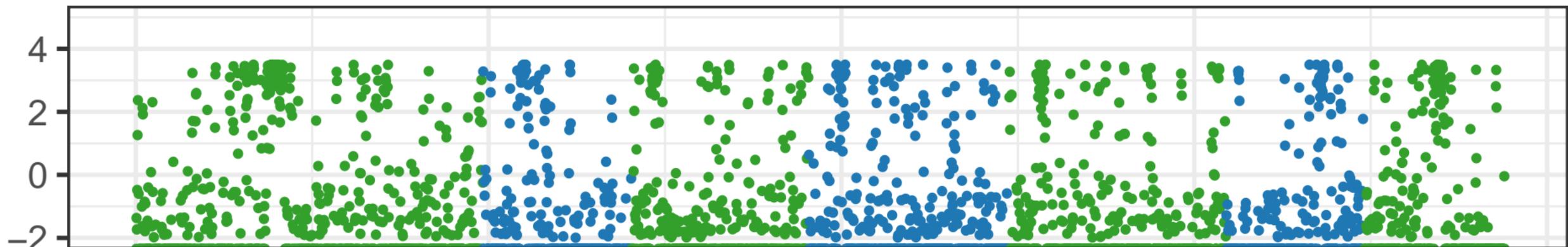
RIP index  $\log_2(\text{obs}/\text{exp})$



window

N.tet\_L7\_9046

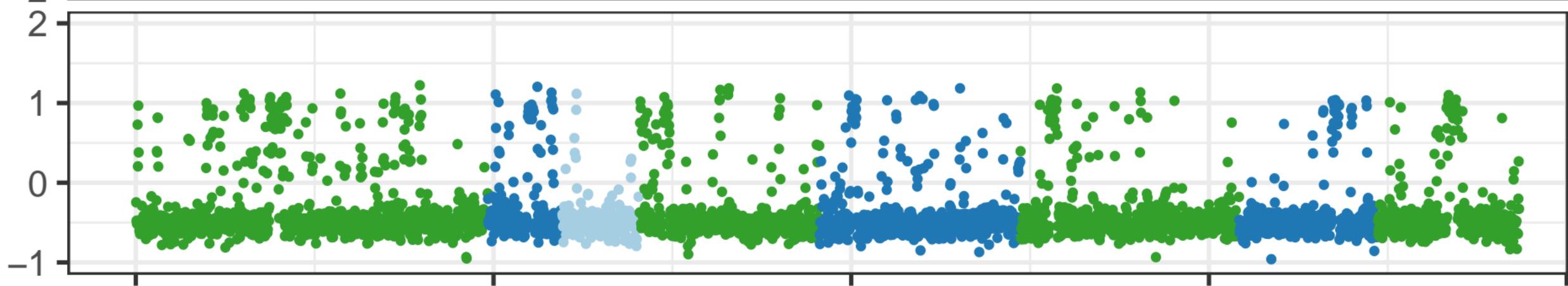
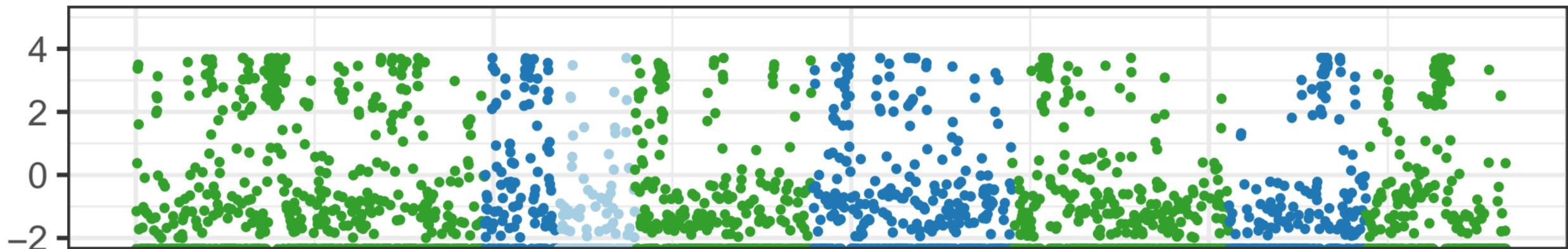
RIP index  $\log_2(\text{obs}/\text{exp})$



window

N.tet\_L8\_2503

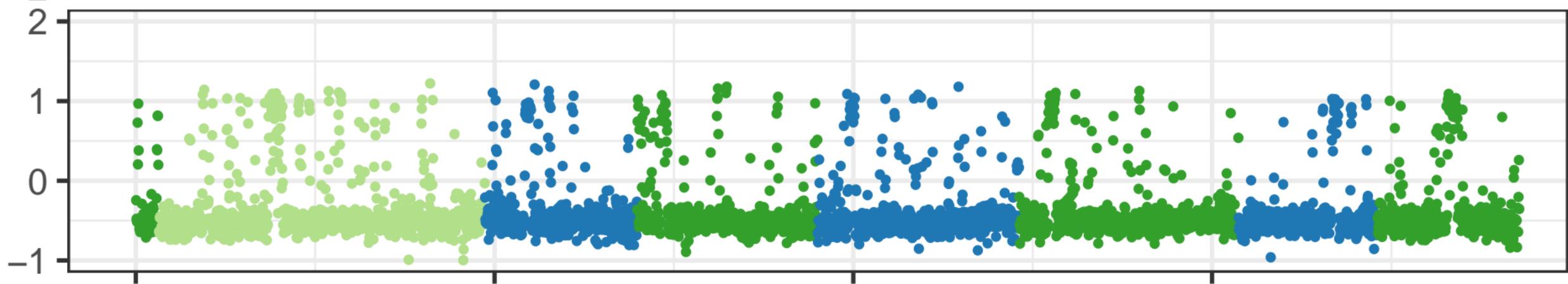
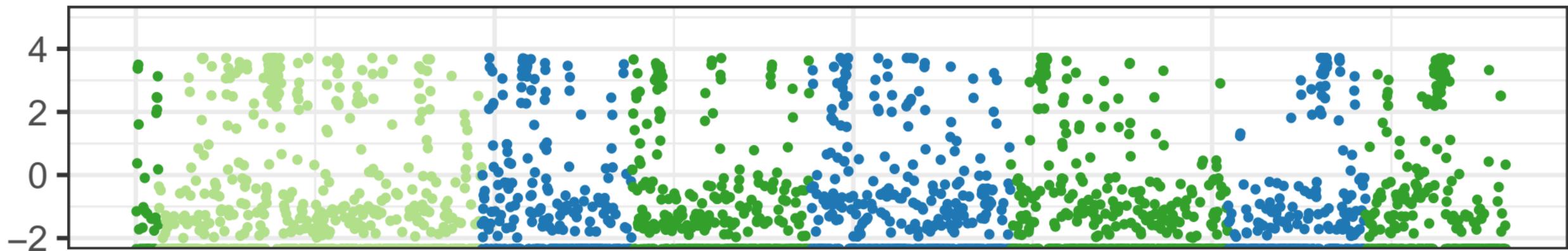
RIP index  $\log_2(\text{obs}/\text{exp})$



window

N.tet\_L8\_2504

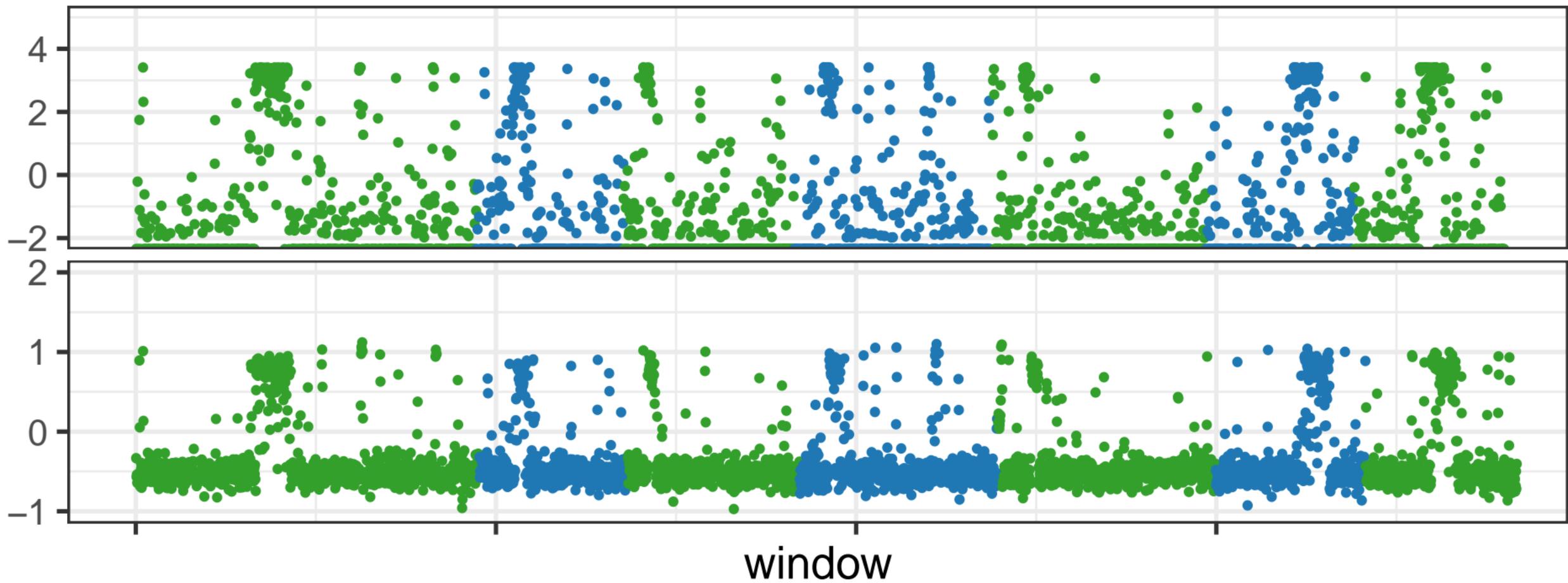
RIP index  $\log_2(\text{obs}/\text{exp})$



window

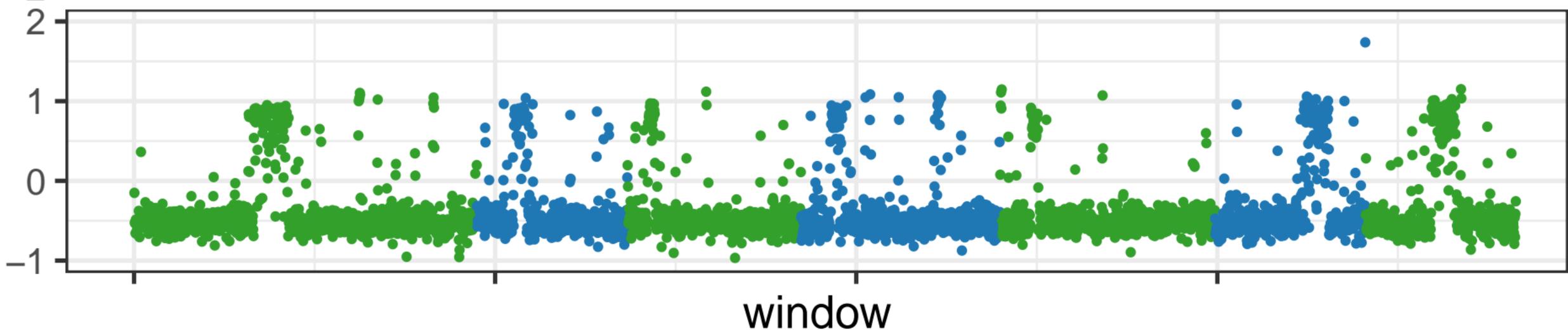
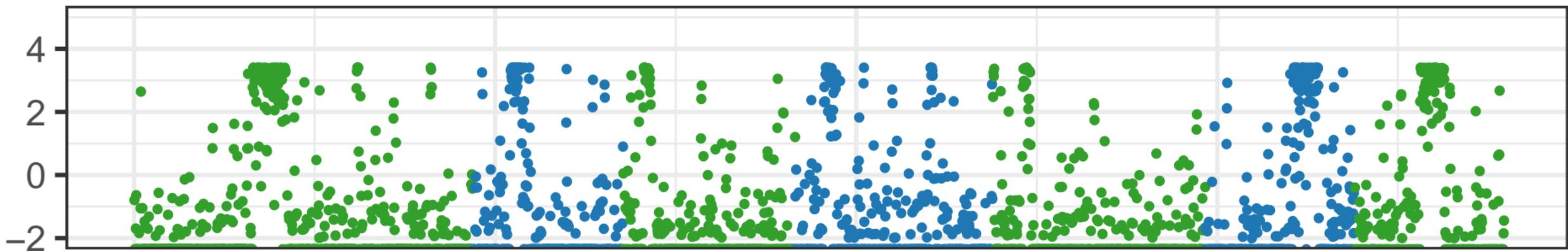
N.sit\_5940

RIP index  $\log_2(\text{obs}/\text{exp})$



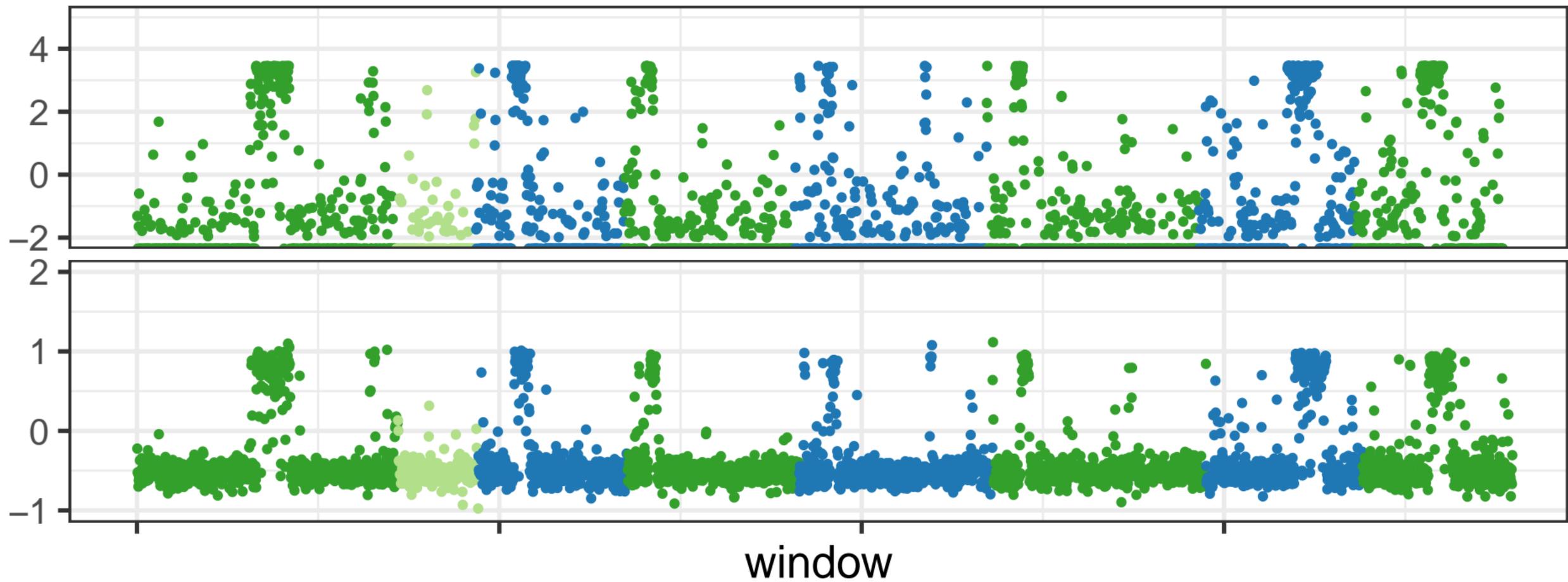
N.sit\_5941

RIP index  $\log_2(\text{obs}/\text{exp})$



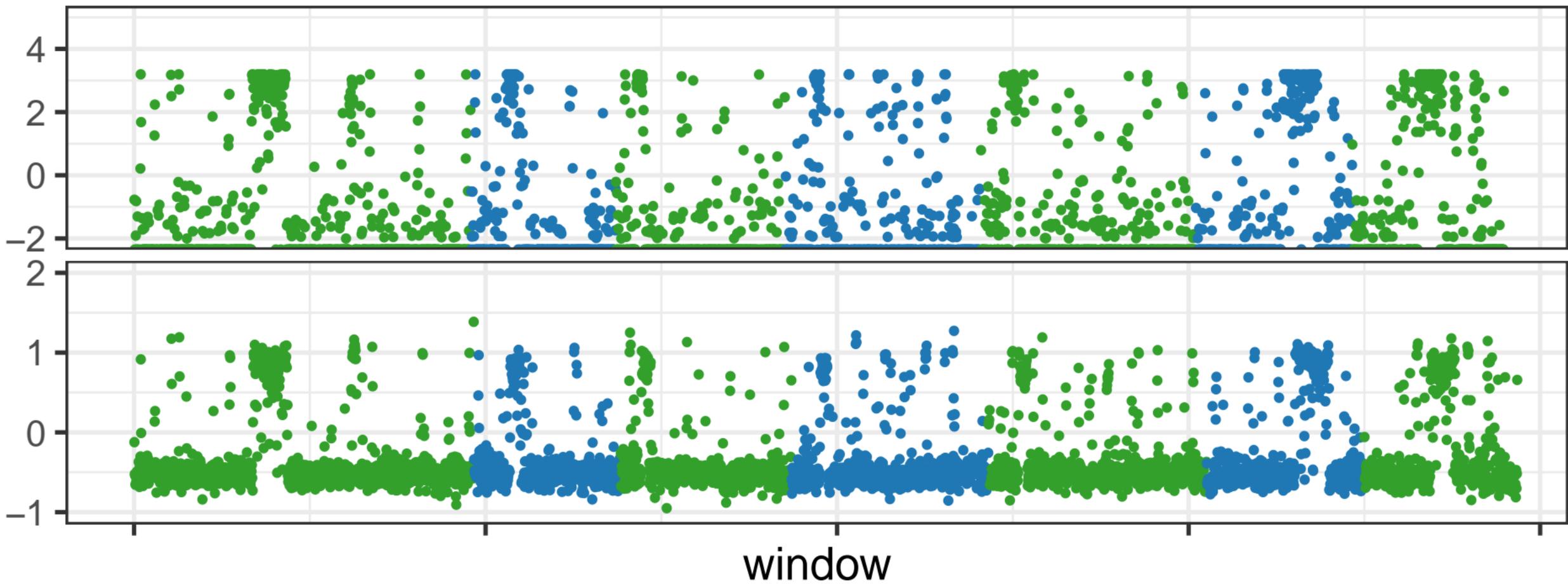
N.sit\_W1426

RIP index  $\log_2(\text{obs}/\text{exp})$



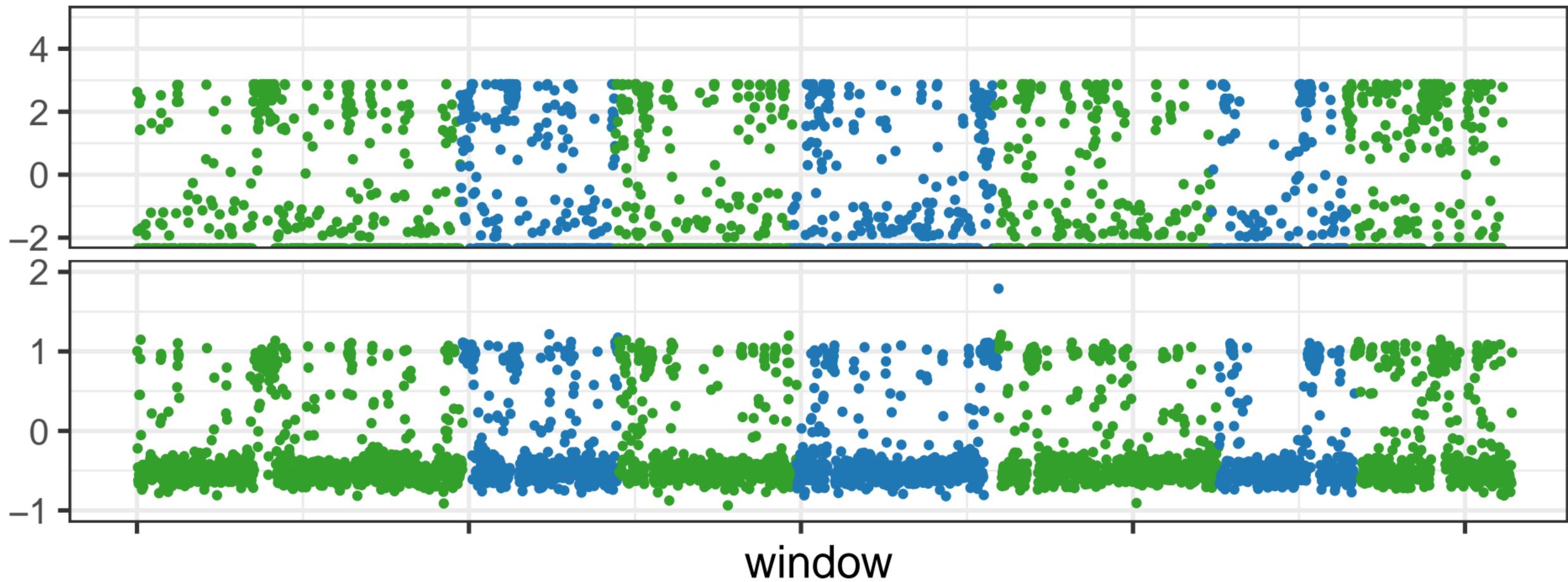
N.sit\_W1434

RIP index  $\log_2(\text{obs}/\text{exp})$



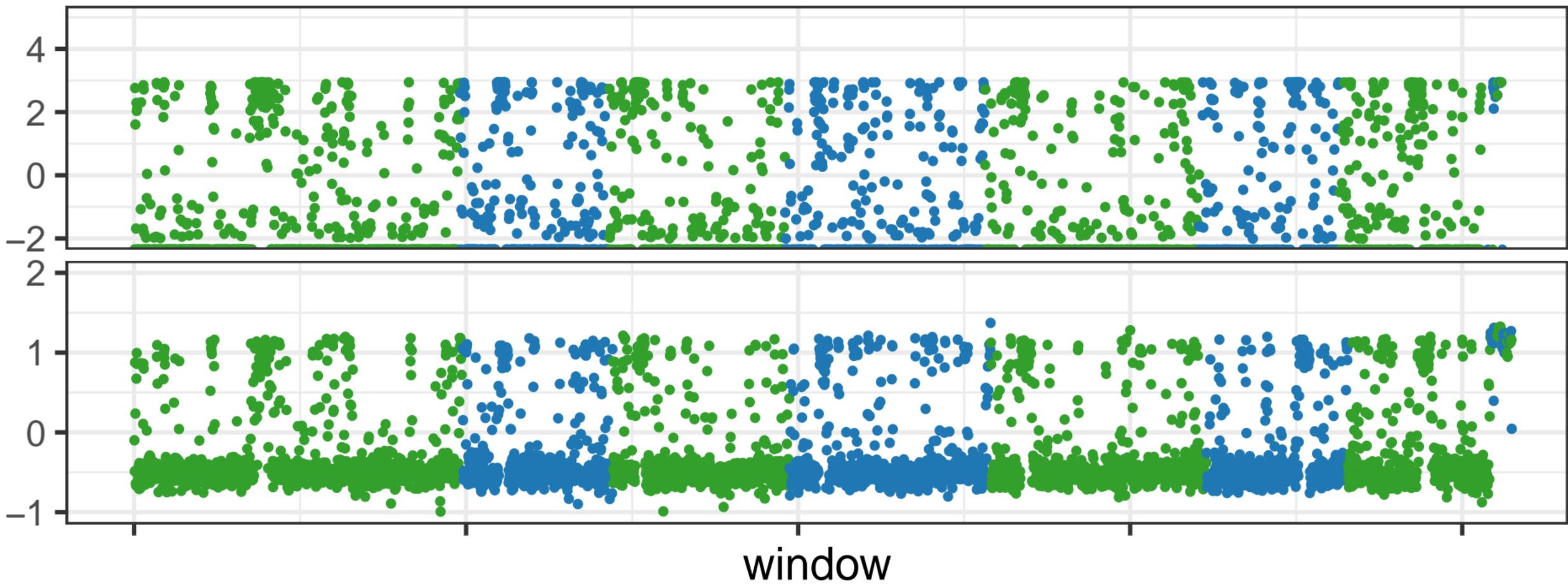
N.his\_8817

RIP index  $\log_2(\text{obs}/\text{exp})$



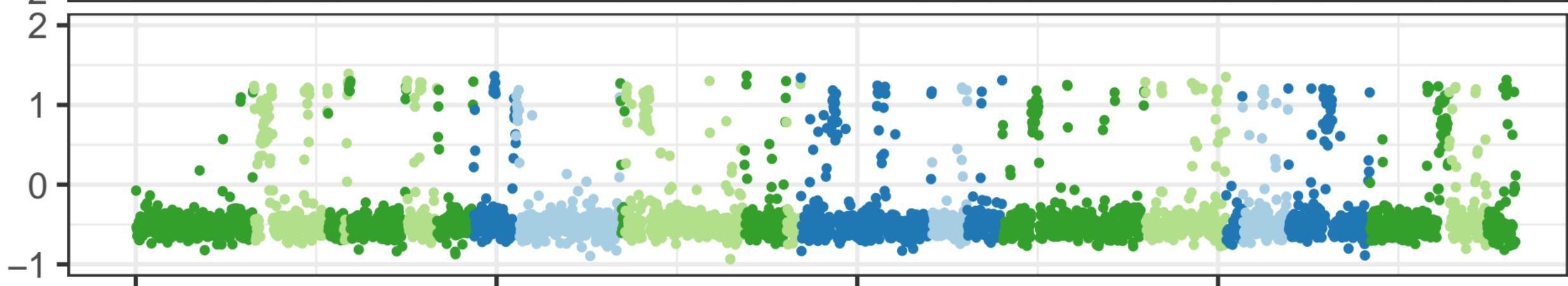
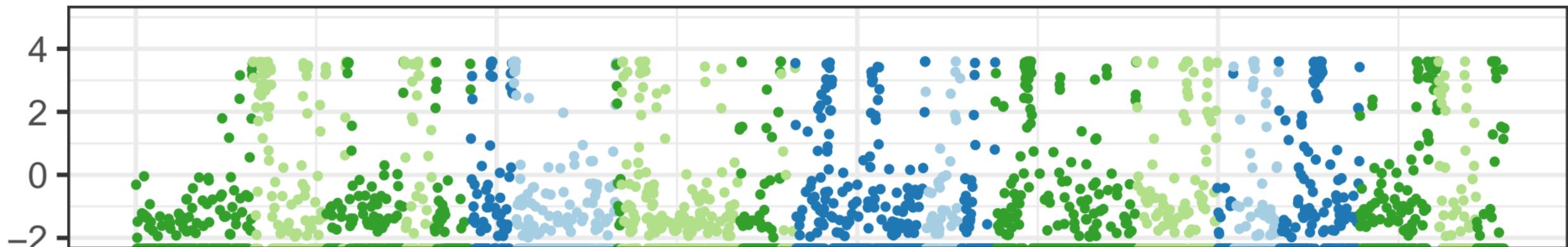
N.cra\_2489

RIP index  $\log_2(\text{obs}/\text{exp})$



N.int\_8767

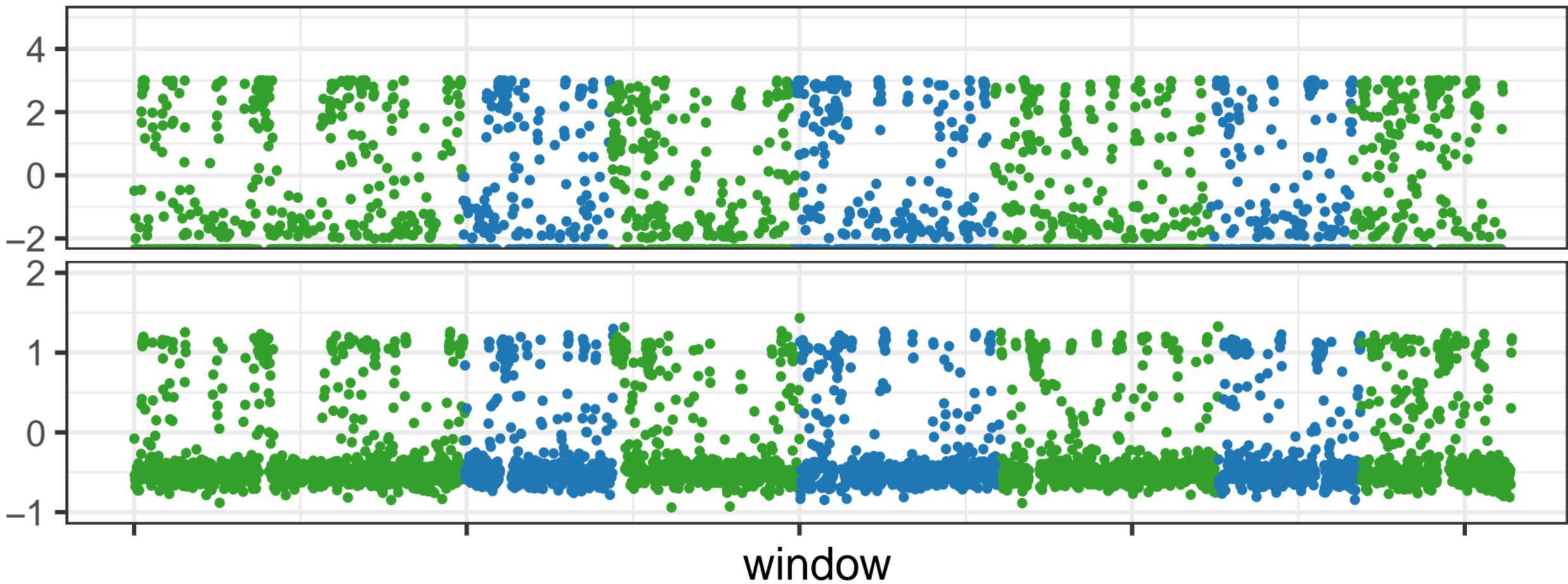
RIP index  $\log_2(\text{obs}/\text{exp})$



window

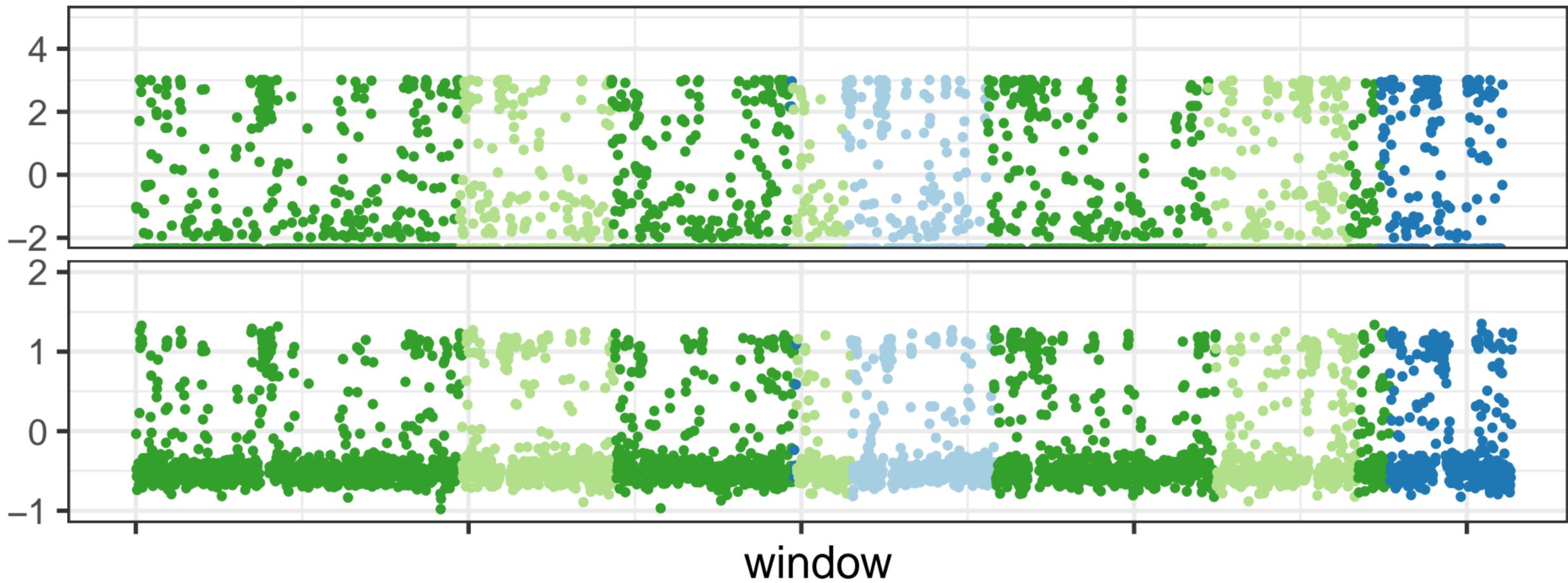
N.int\_8807

RIP index  $\log_2(\text{obs}/\text{exp})$



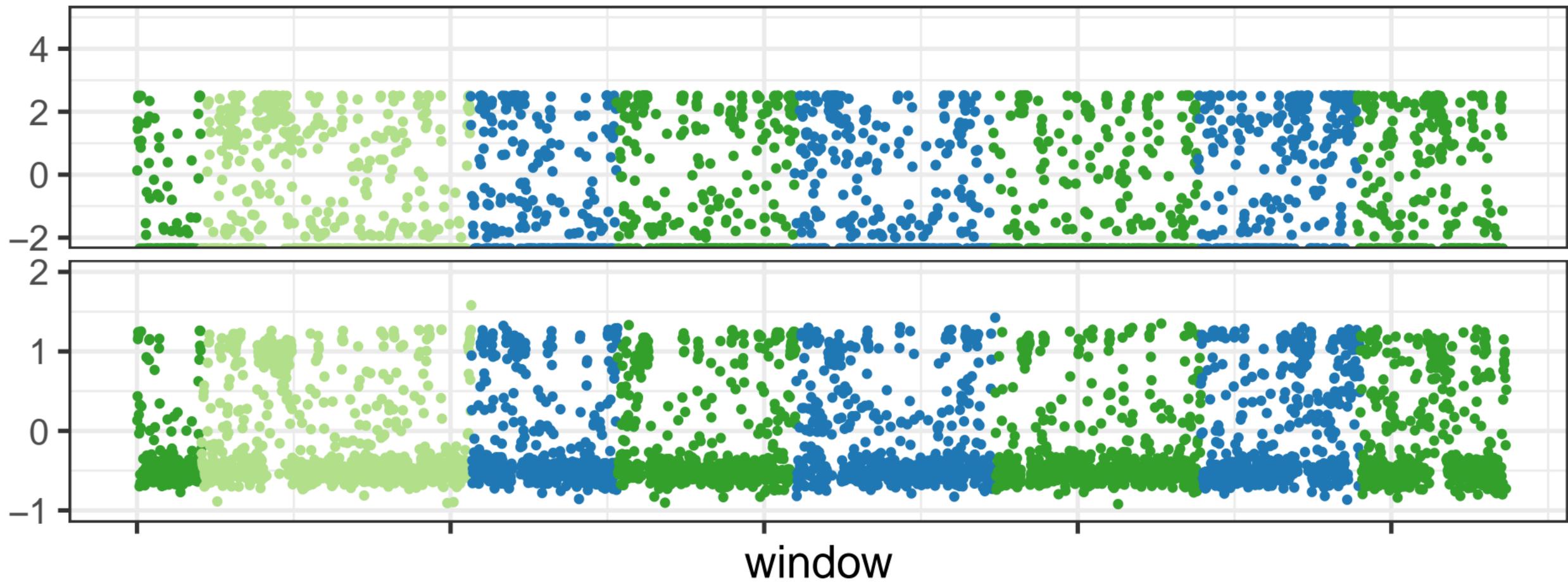
N.int\_8761

RIP index  $\log_2(\text{obs}/\text{exp})$



N.met\_10397

RIP index  $\log_2(\text{obs}/\text{exp})$



N.dis\_8579

RIP index  $\log_2(\text{obs}/\text{exp})$

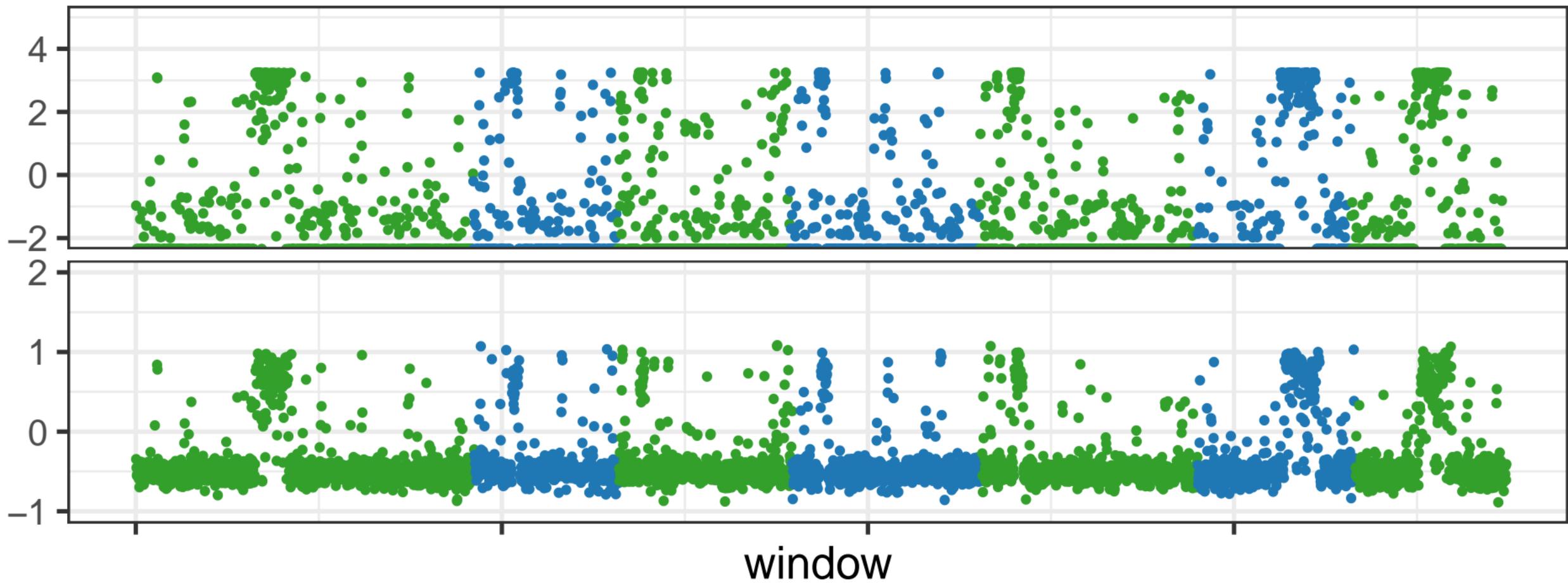
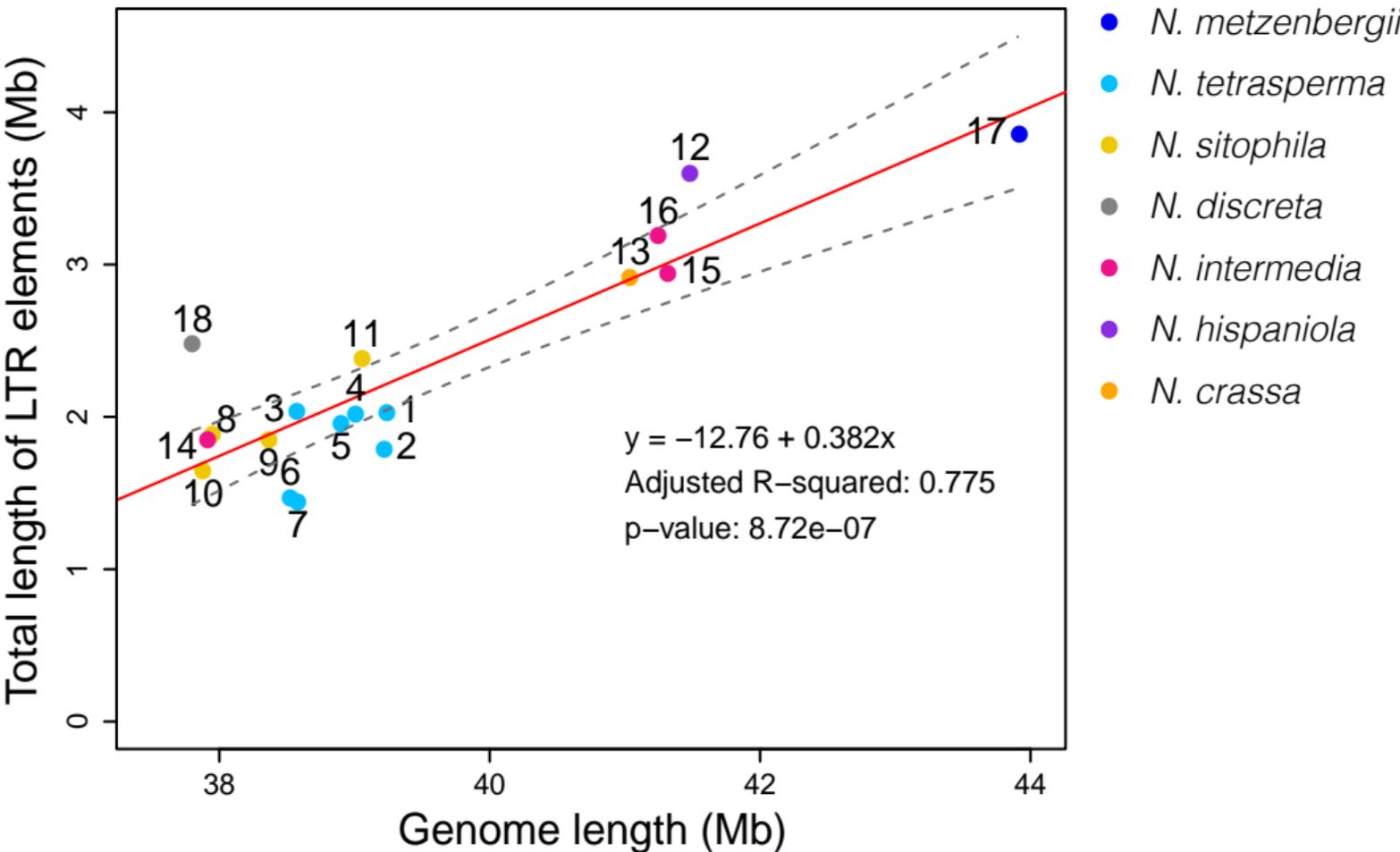
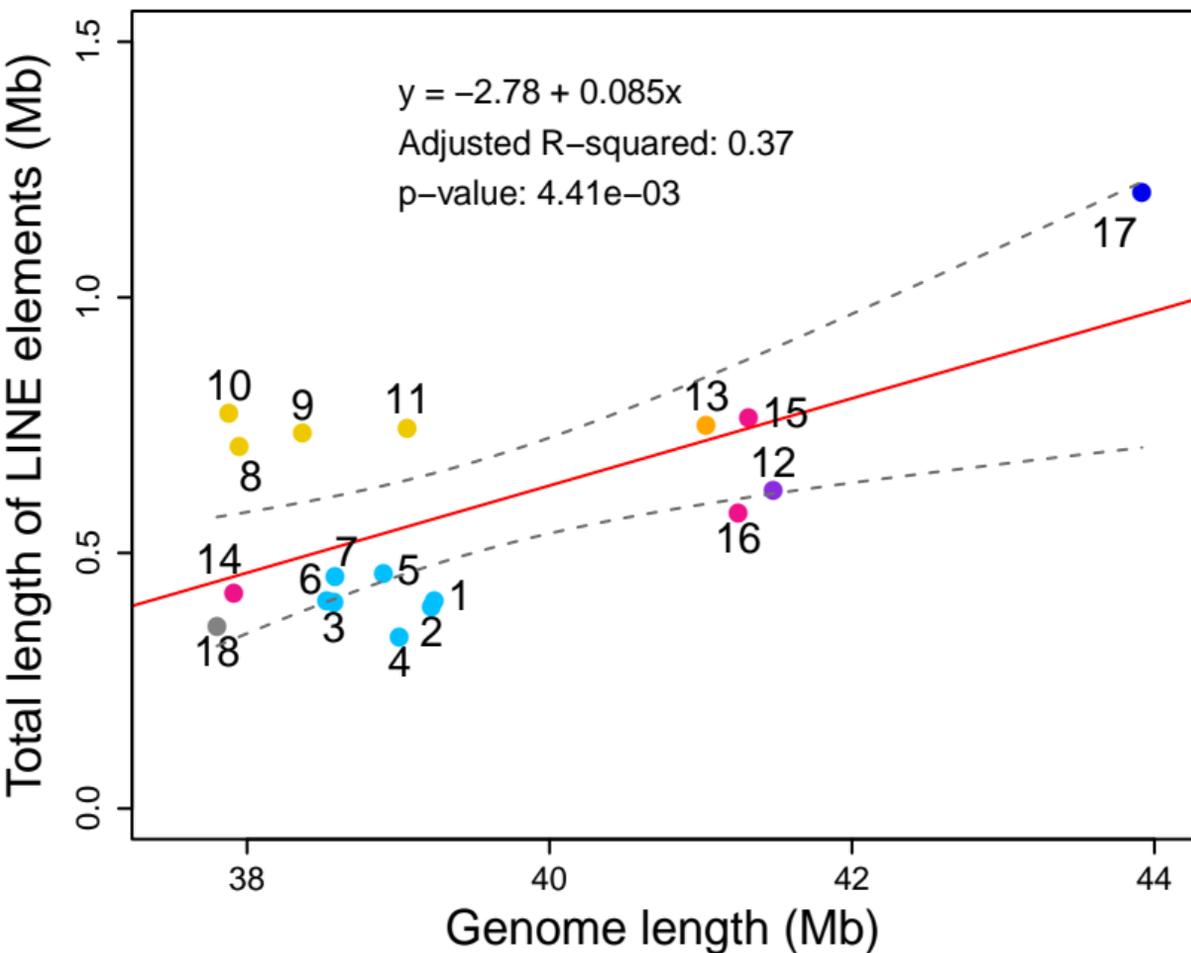


Figure S3

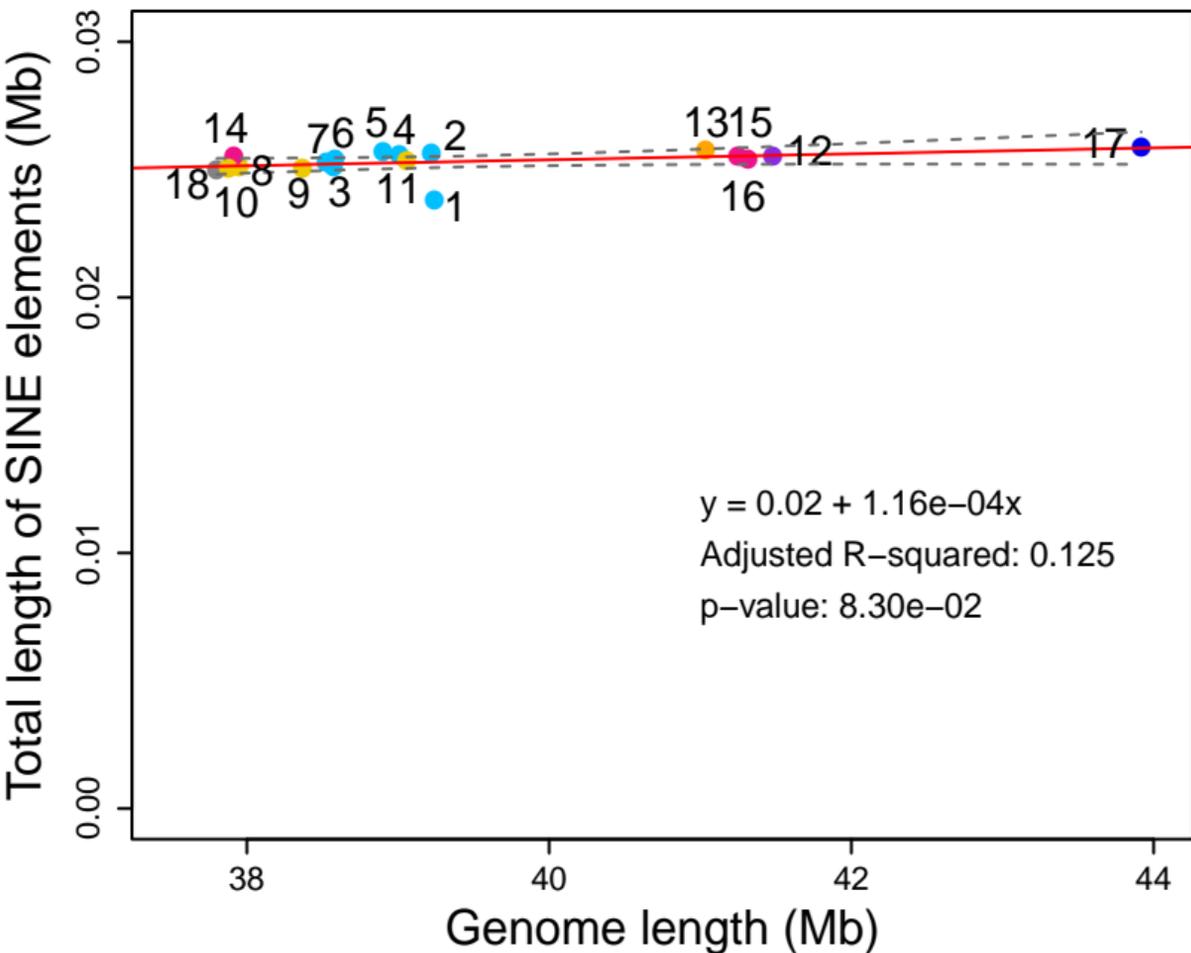
# LTR elements



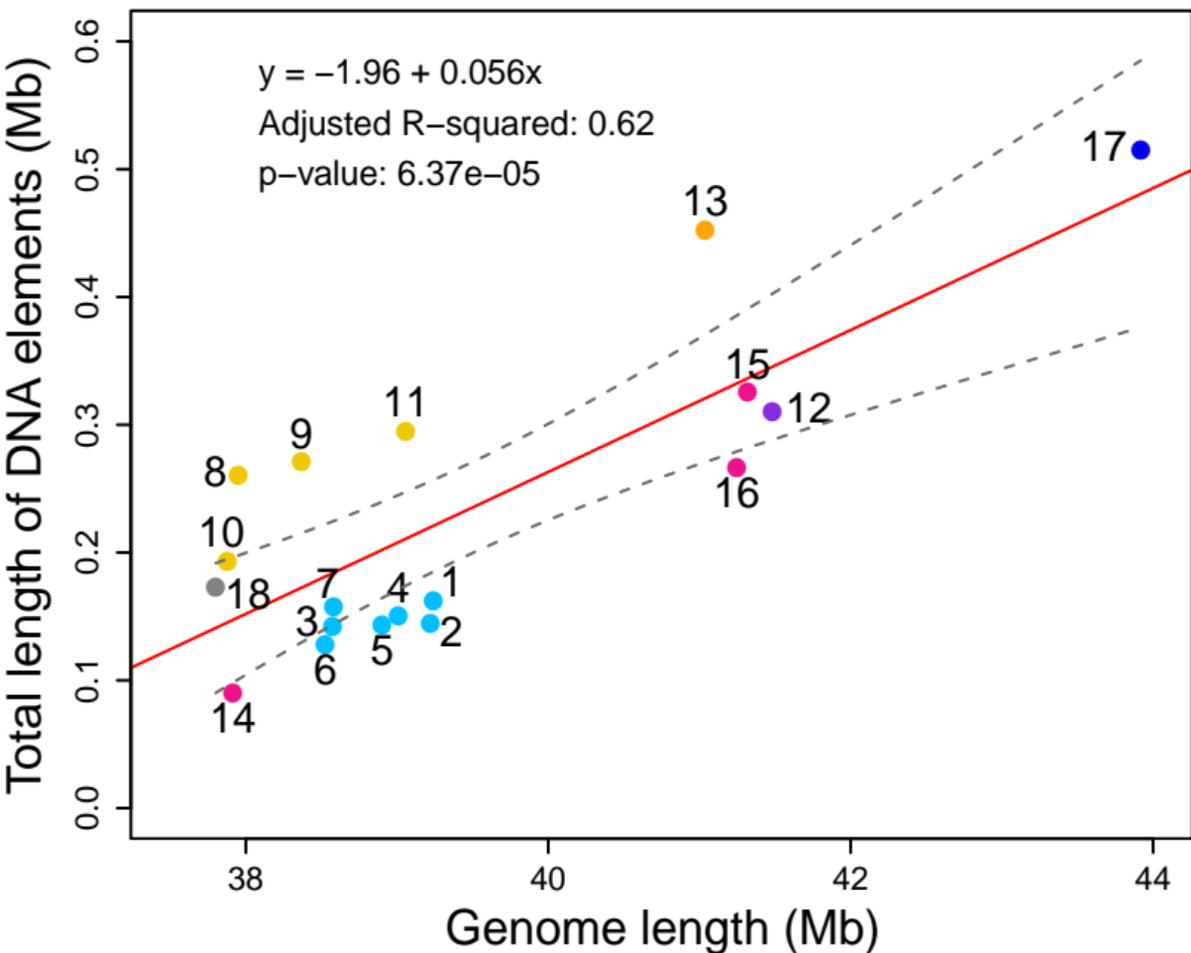
# LINE elements



# SINE elements

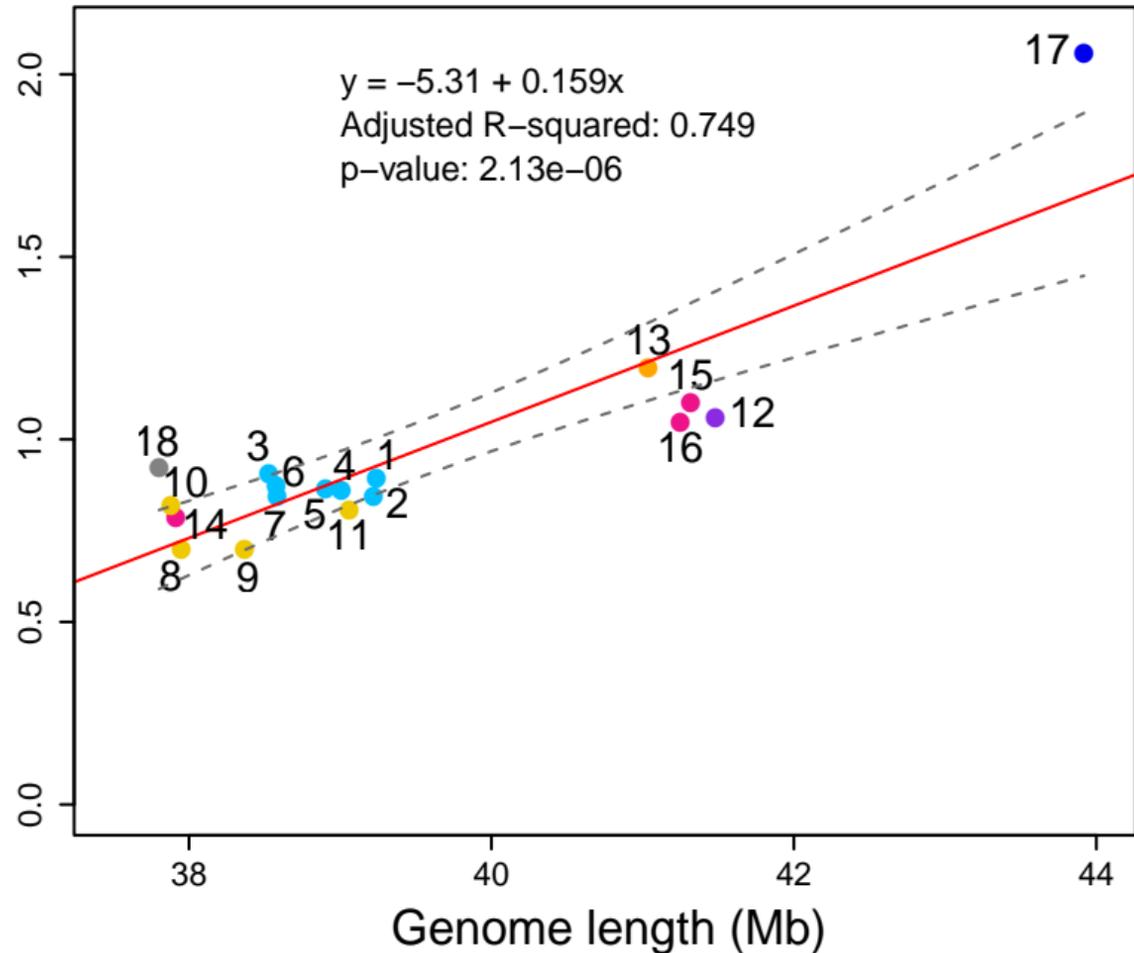


# DNA elements



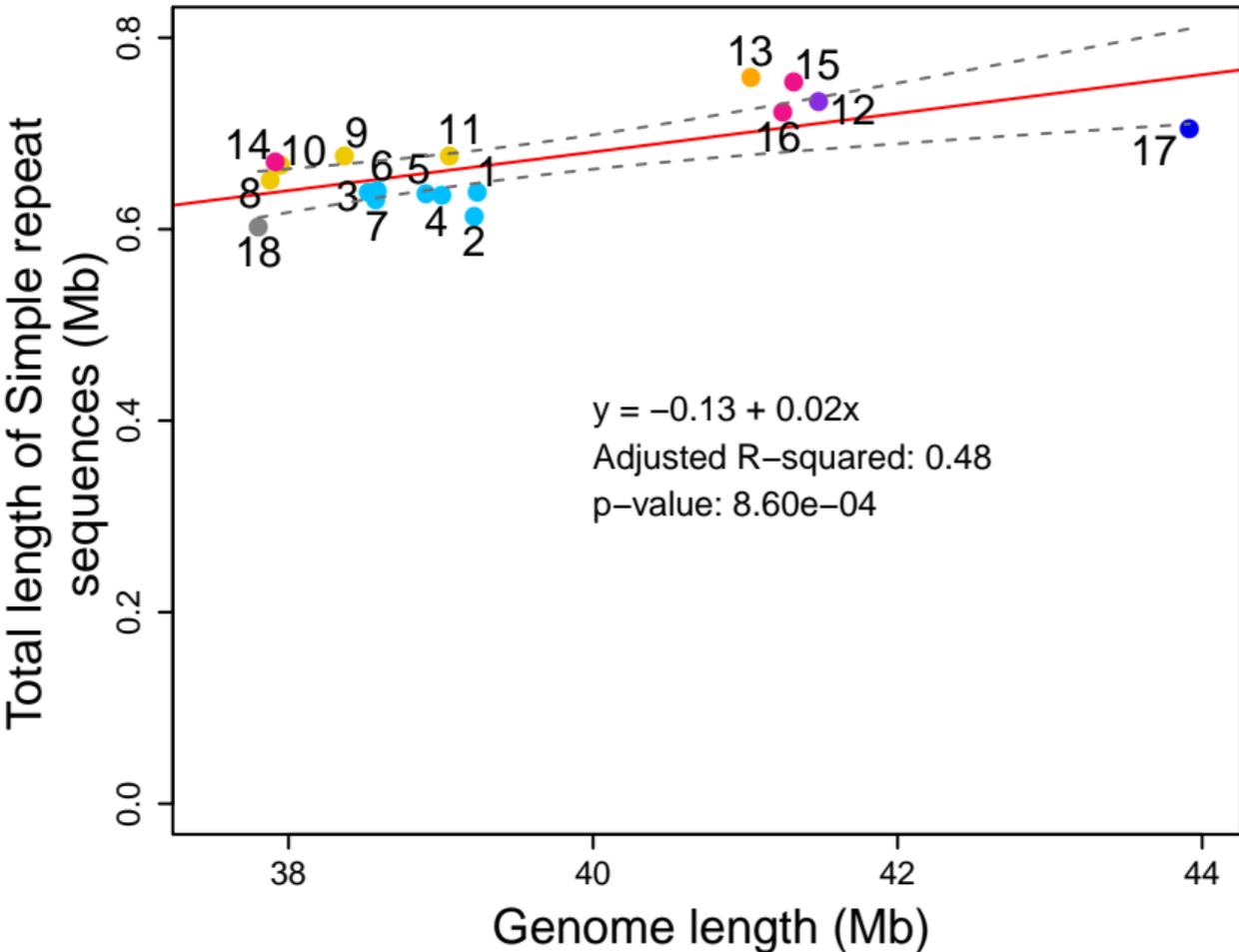
# Unclassified elements

Total length of Unclassified elements (Mb)



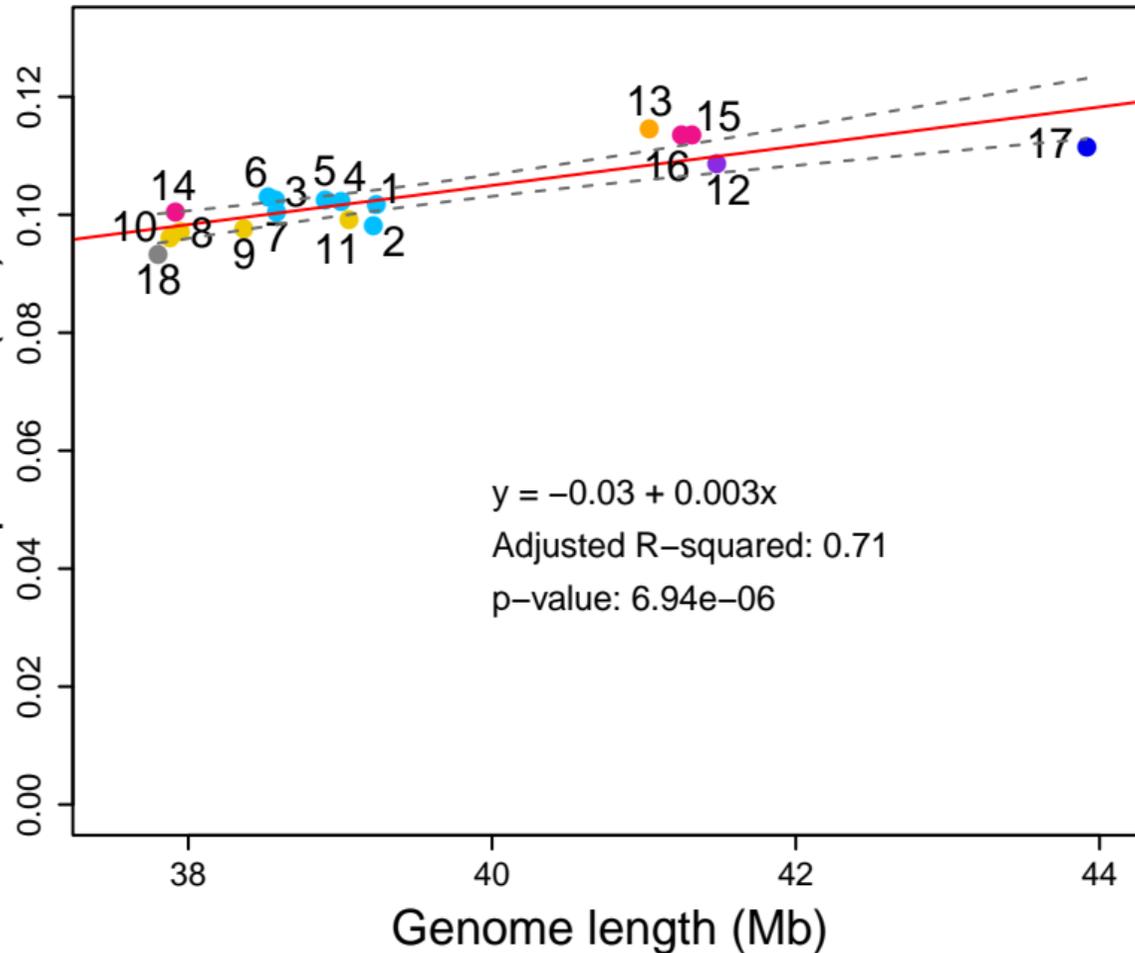
- *N. metzenbergii*
- *N. tetrasperma*
- *N. sitophila*
- *N. discreta*
- *N. intermedia*
- *N. hispaniola*
- *N. crassa*

# Simple repeat sequences



# Low complexity sequences

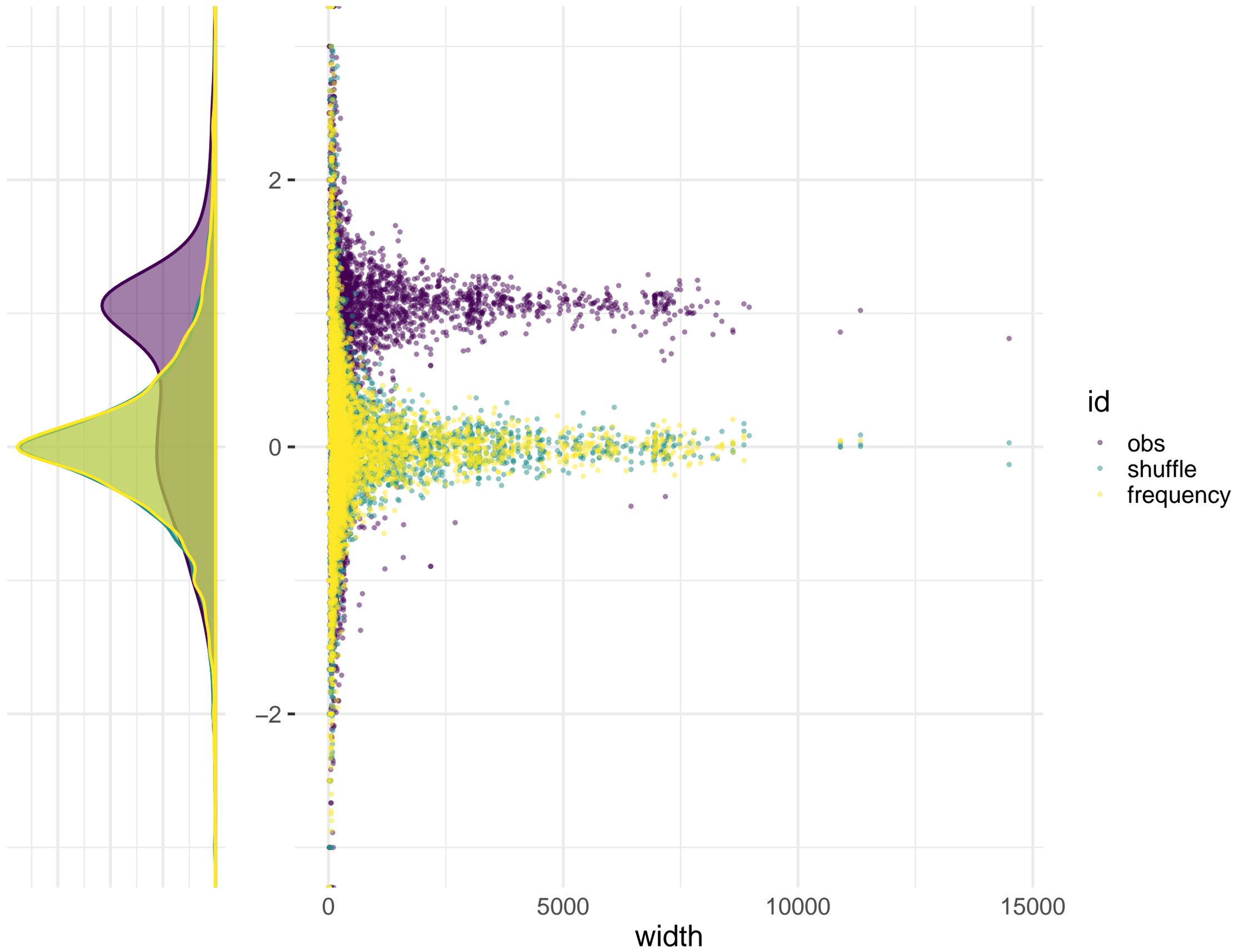
Total length of Low complexity sequences (Mb)



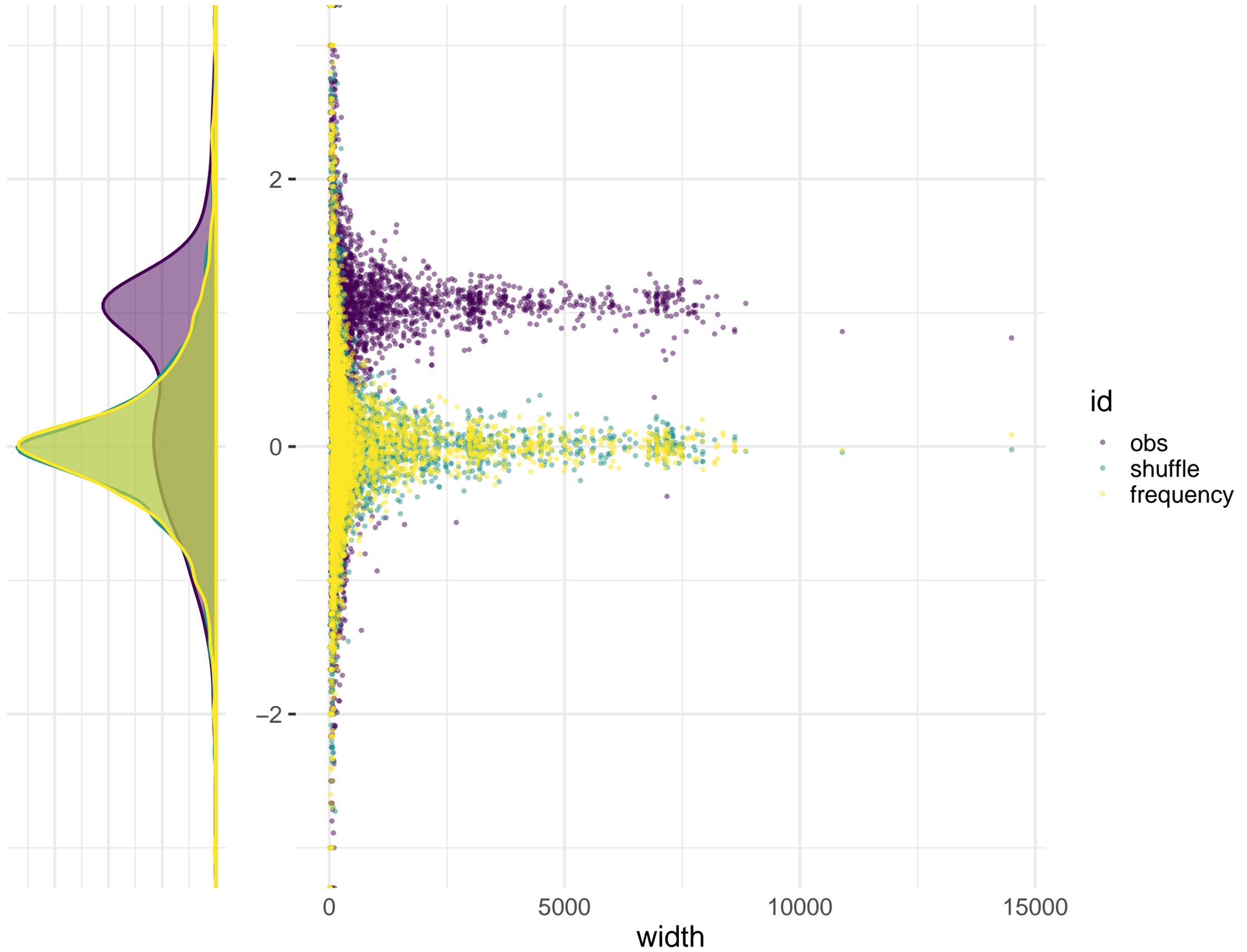
- *N. metzenbergii*
- *N. tetrasperma*
- *N. sitophila*
- *N. discreta*
- *N. intermedia*
- *N. hispaniola*
- *N. crassa*

Figure S4

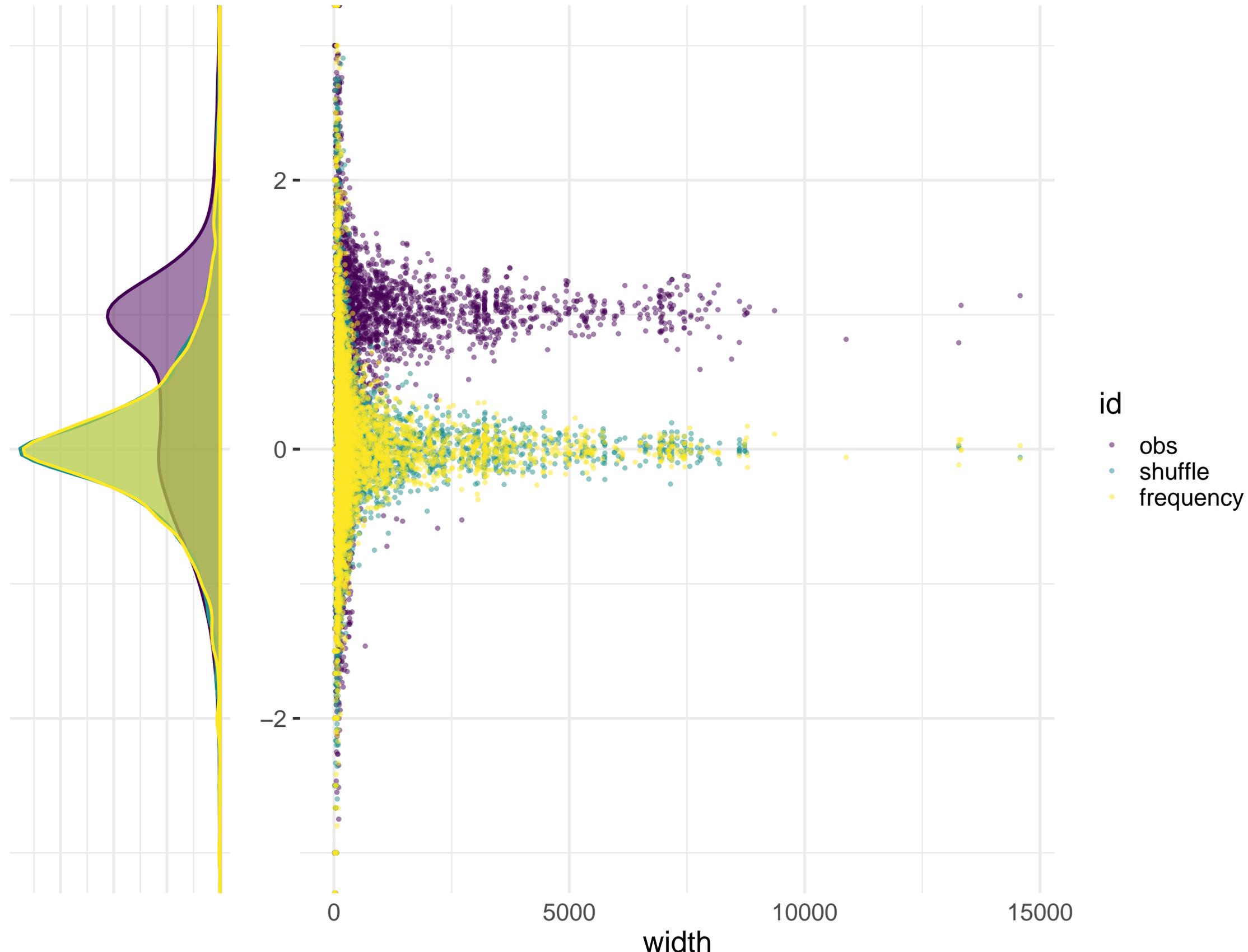
N.tet\_L6\_2508



N.tet\_L6\_2509



N.tet\_L9\_10752



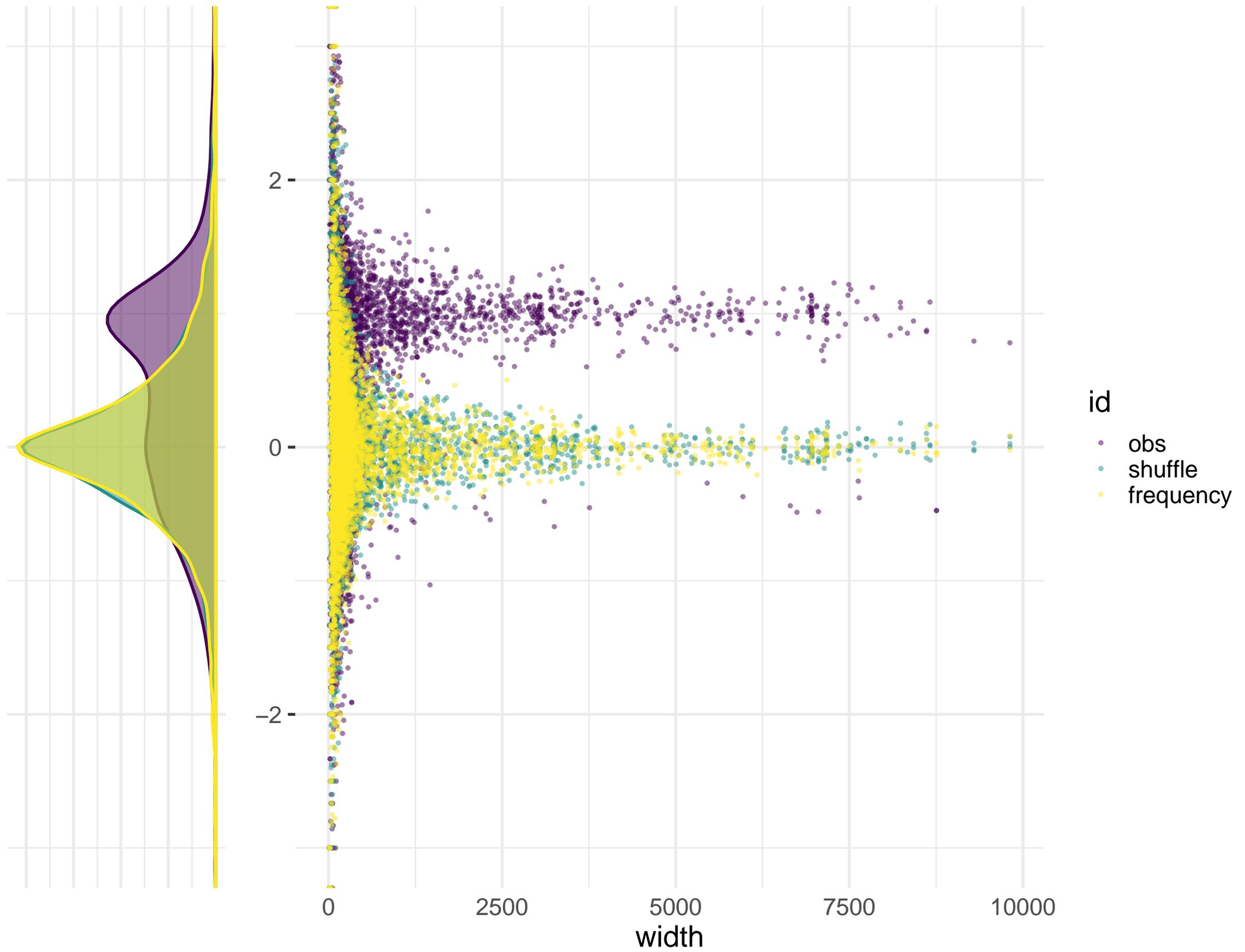
N.tet\_L7\_9045



N.tet\_L7\_9046



N.tet\_L8\_2503



N.tet\_L8\_2504



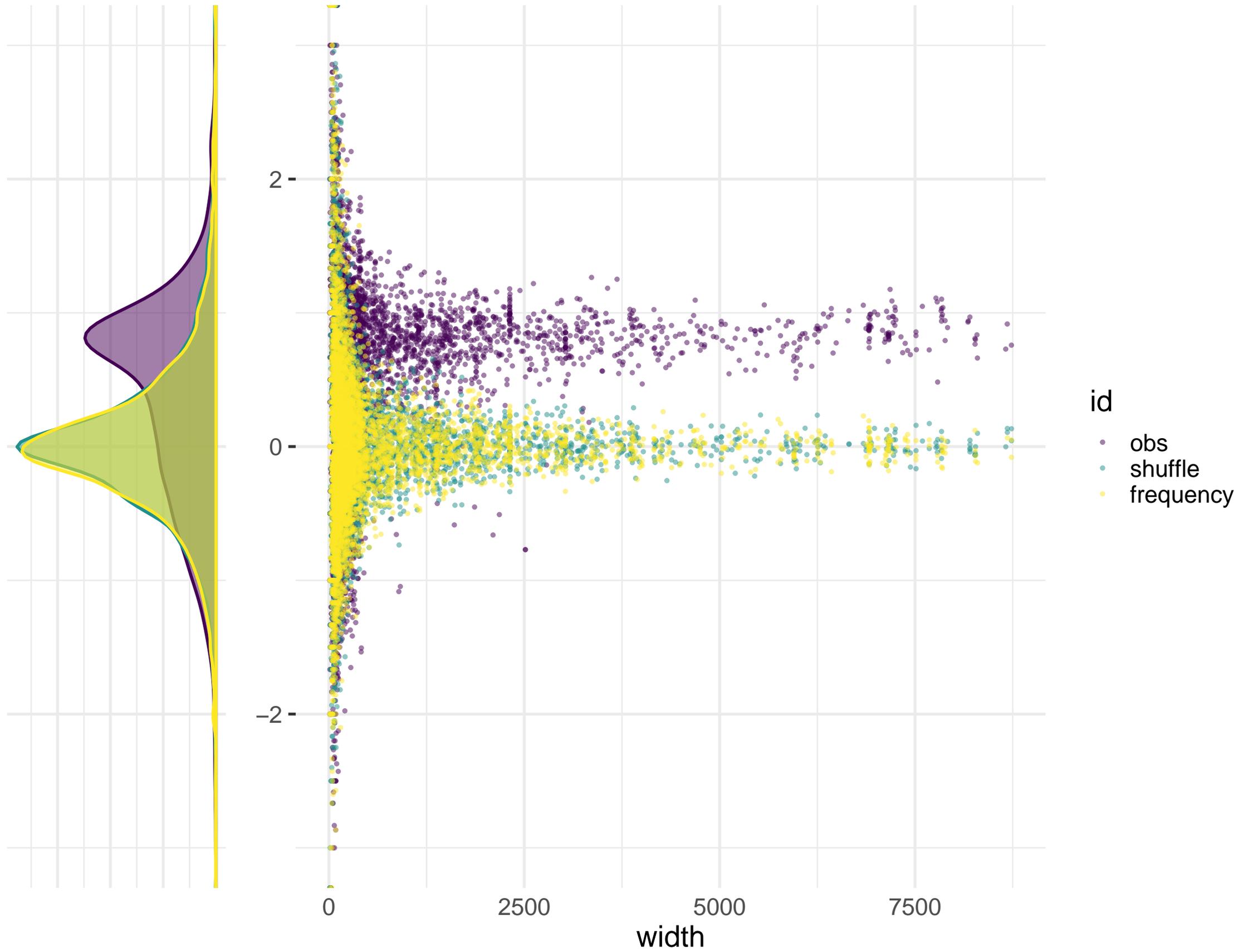
N.sit\_5940



N.sit\_5941



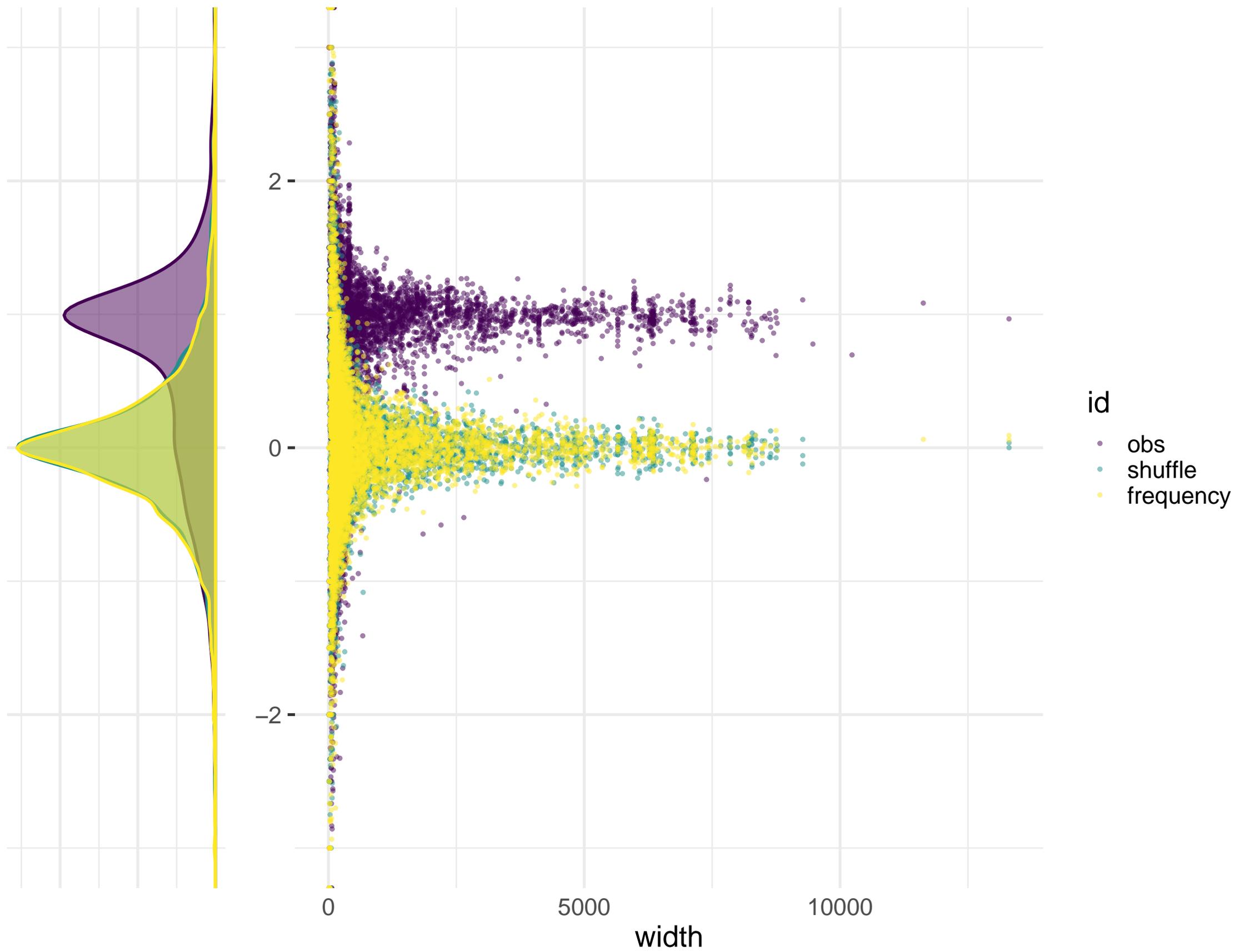
N.sit\_W1426



N.sit\_W1434



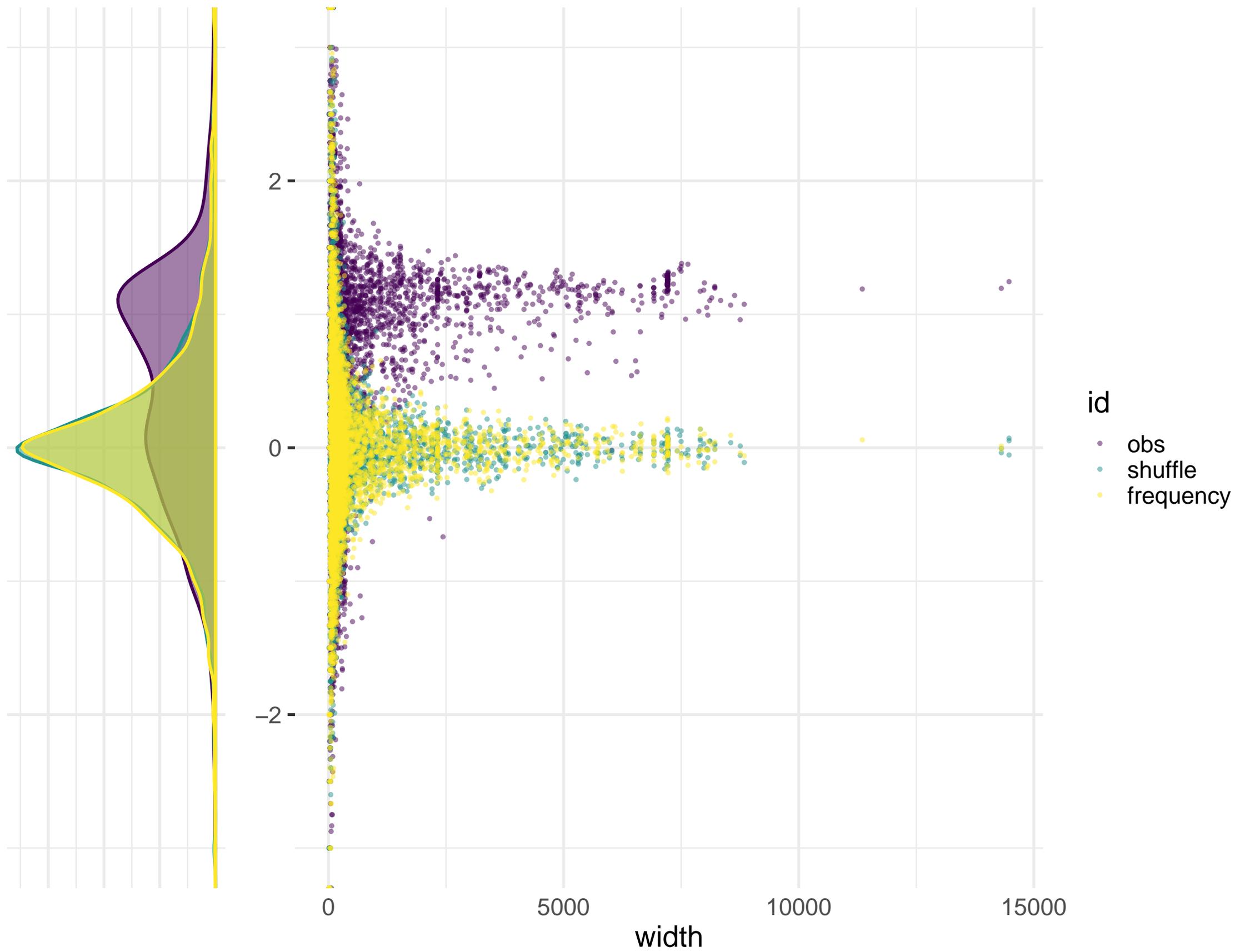
N.his\_8817



N.cra\_2489



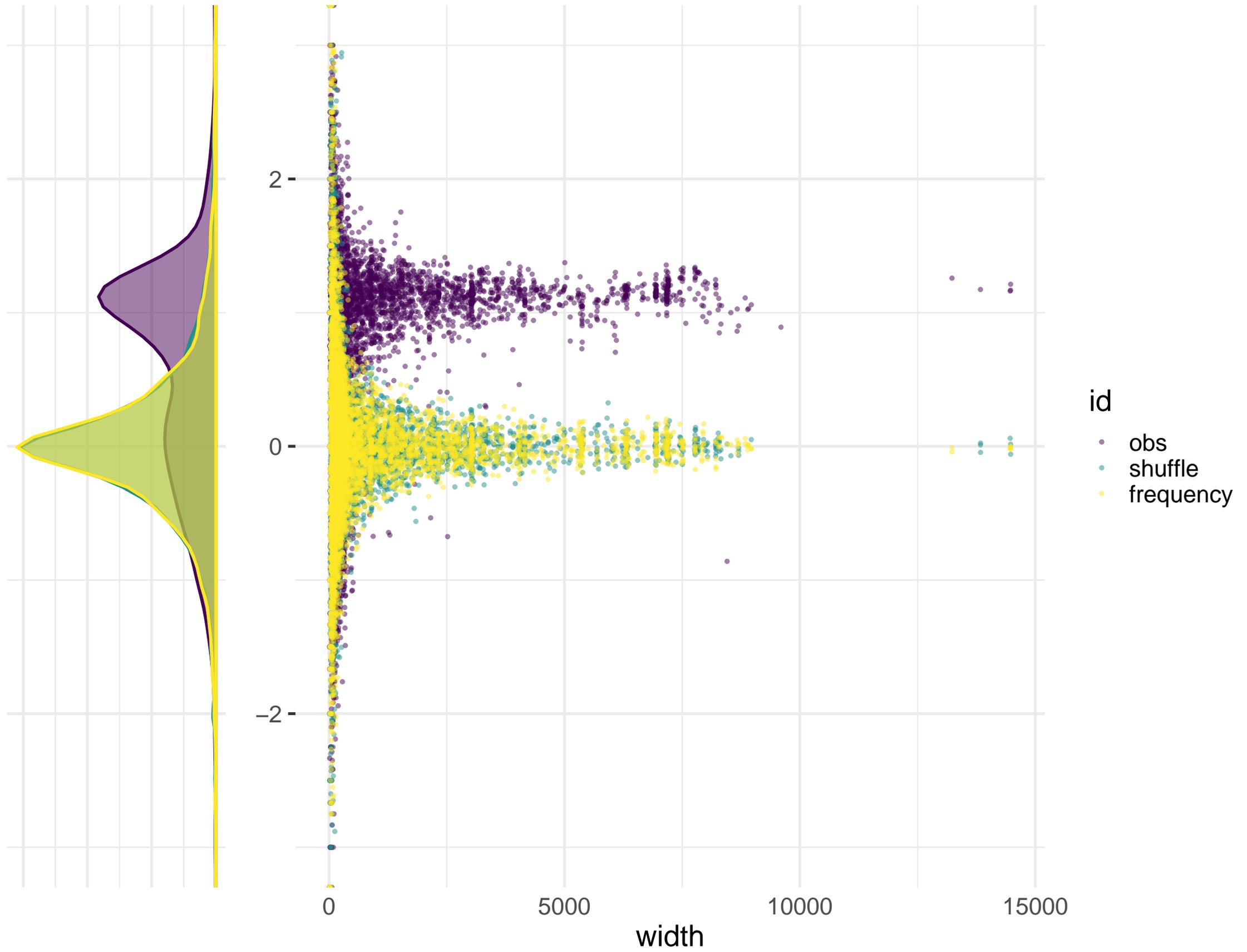
N.int\_8767



N.int\_8807



N.int\_8761



N.met\_10397



N.dis\_8579



Figure S5

N.tet\_L6\_2508

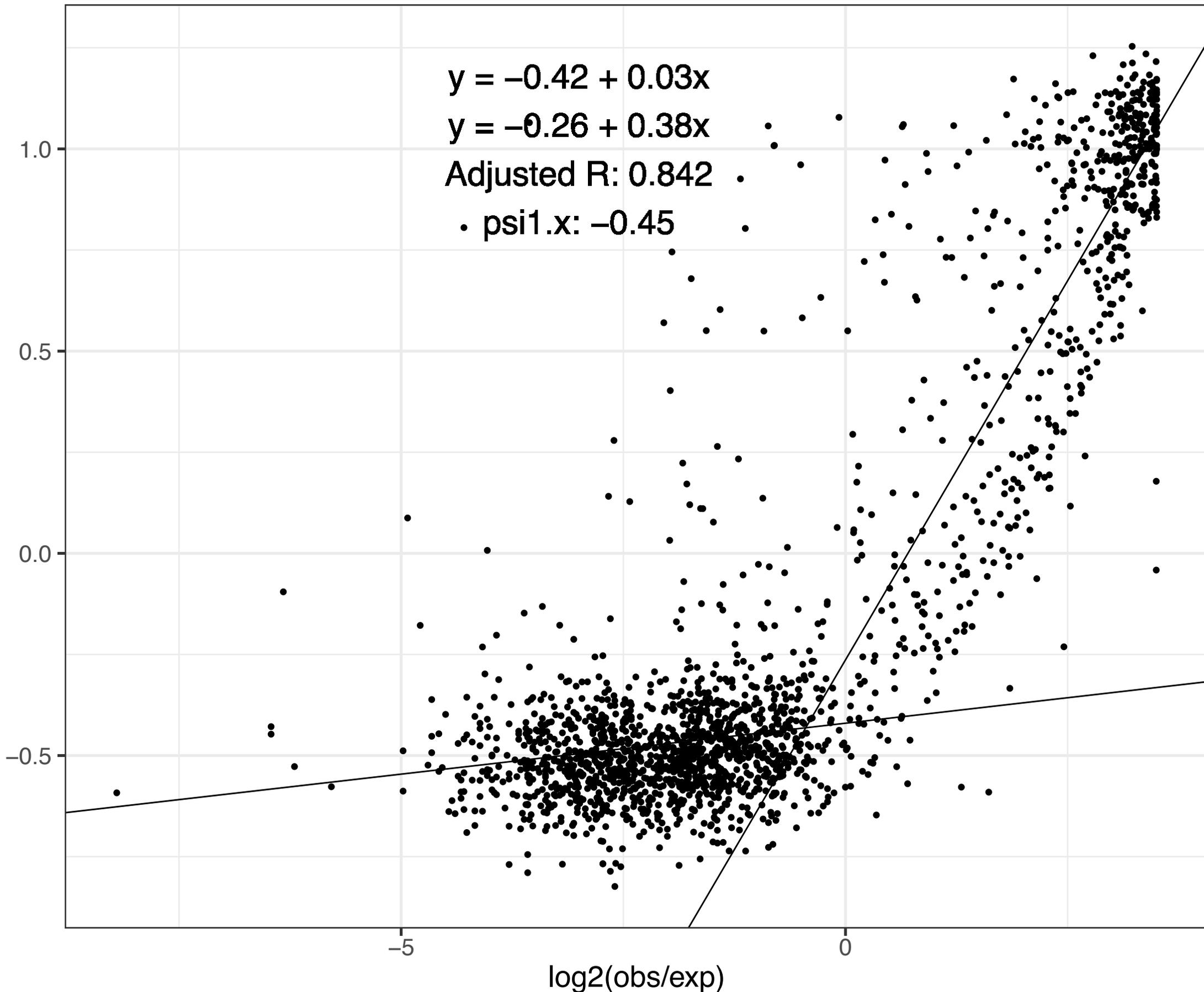
$y = -0.42 + 0.03x$

$y = -0.26 + 0.38x$

Adjusted R: 0.842

• psi1.x: -0.45

RIP index

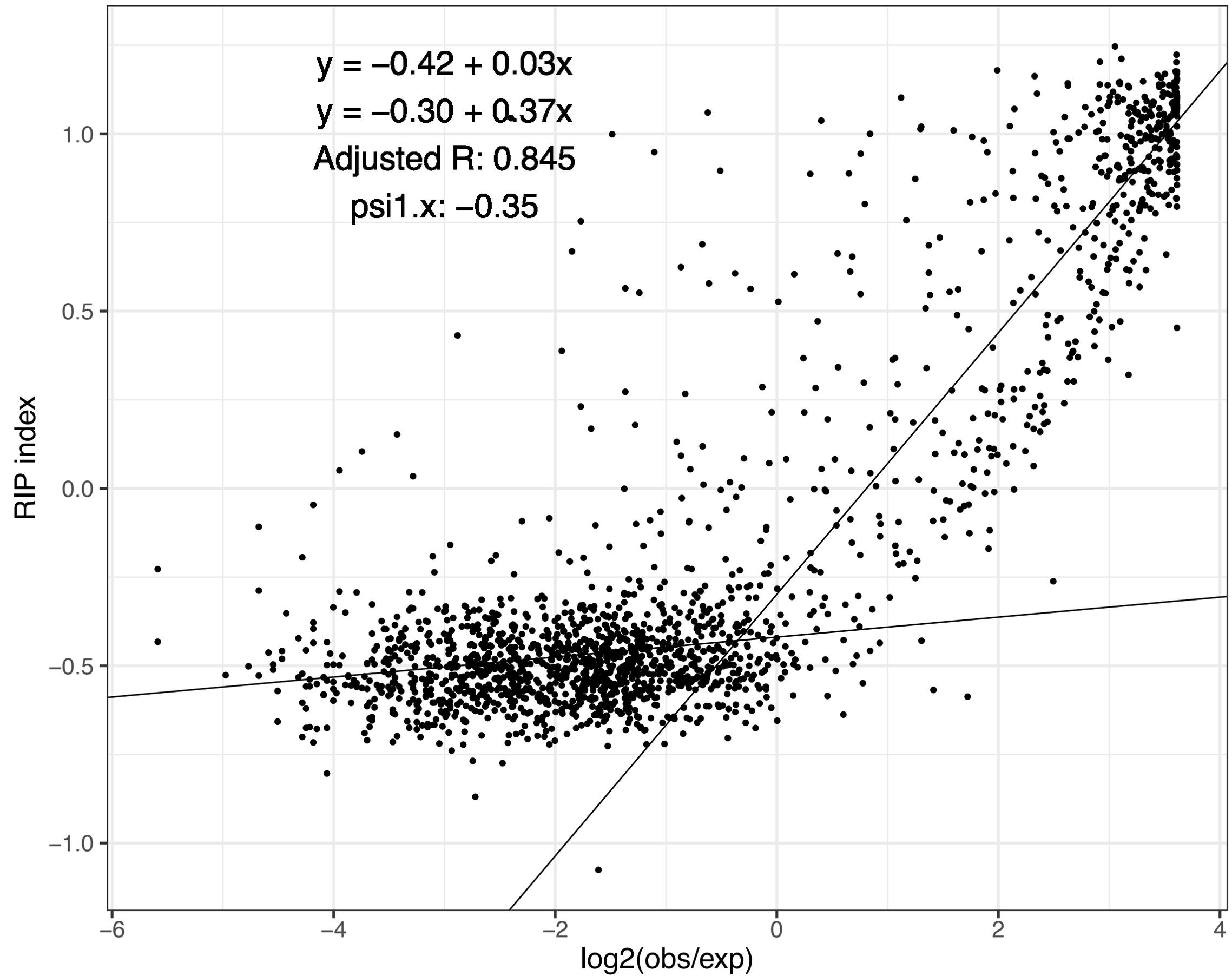


-5

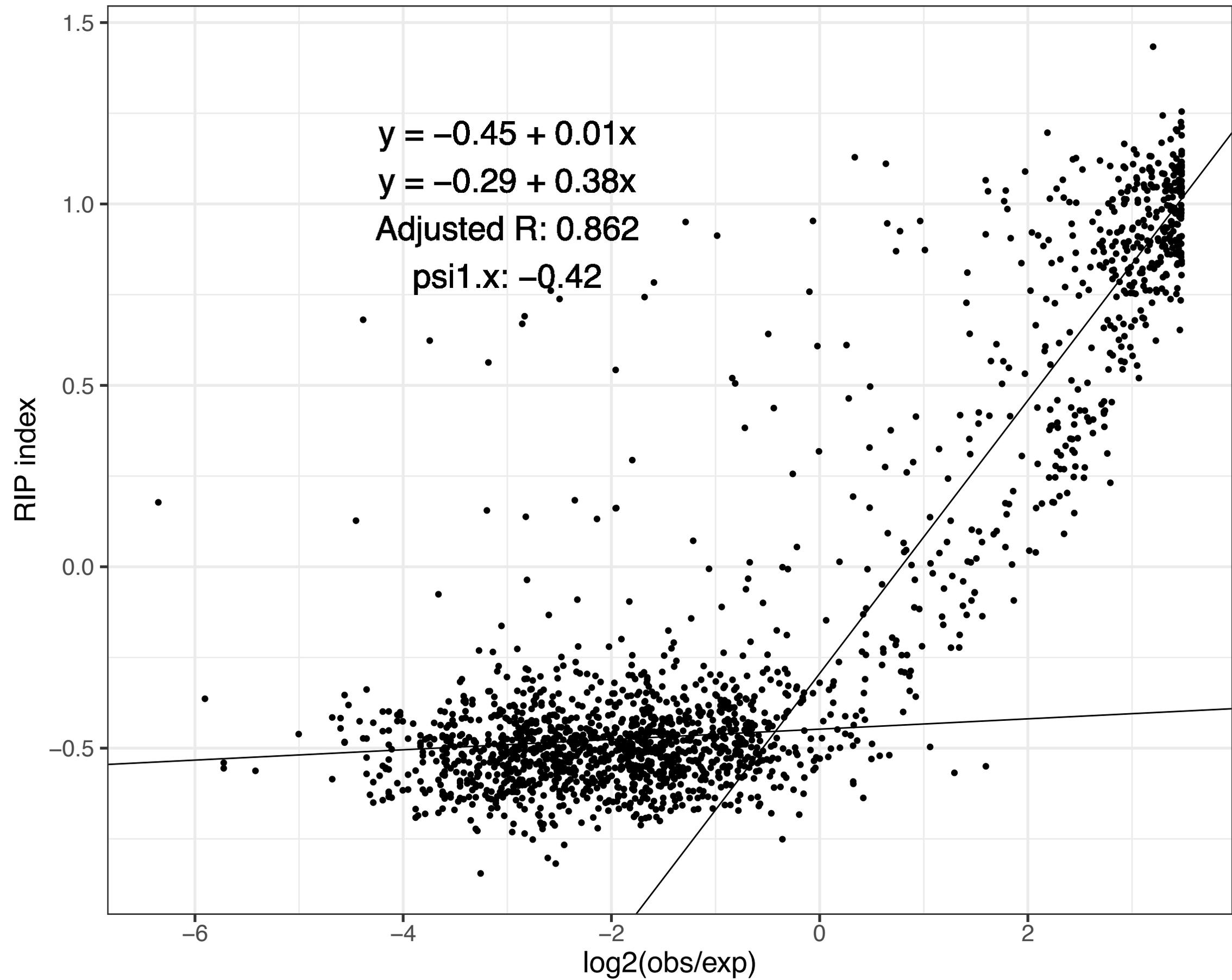
0

log2(obs/exp)

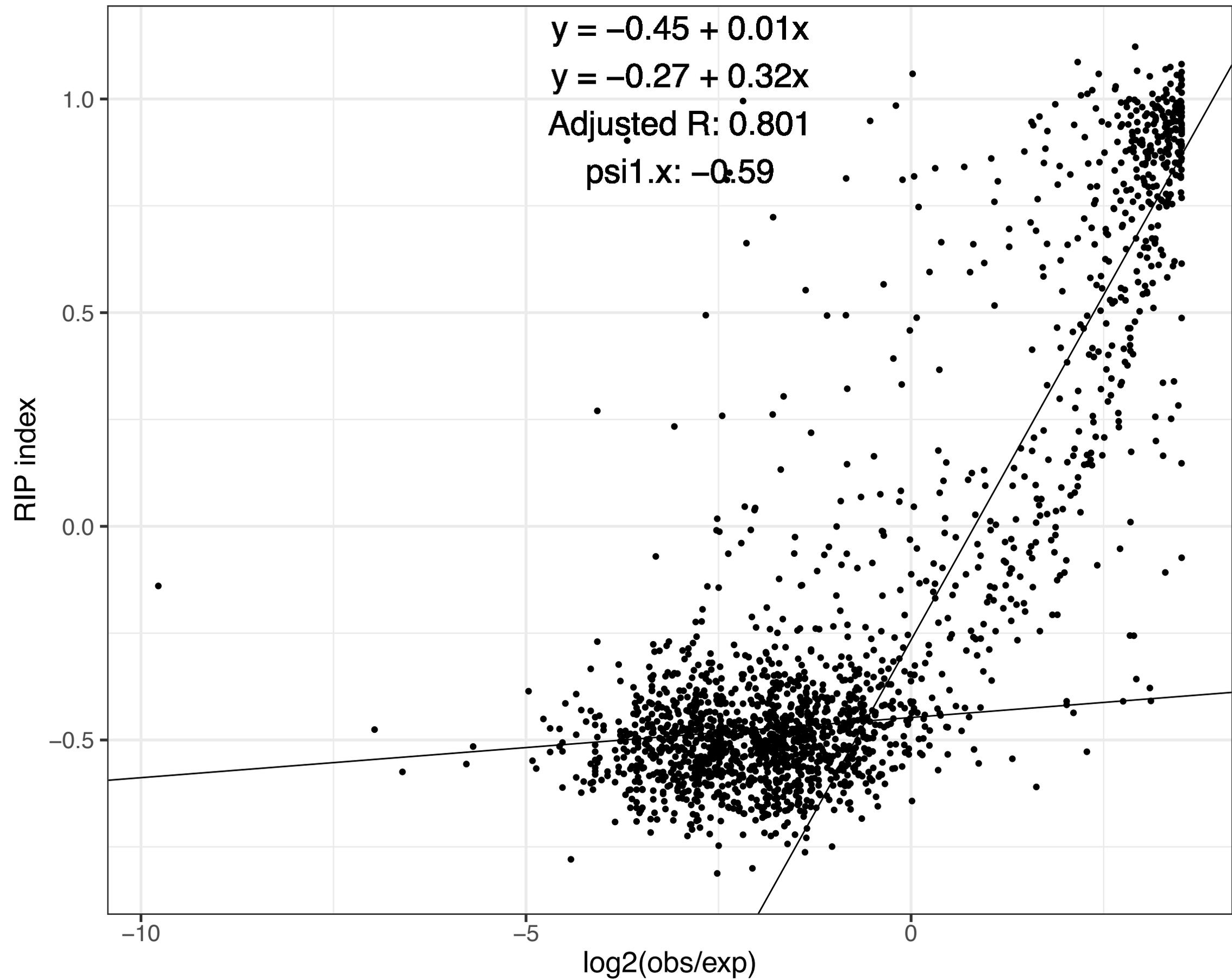
N.tet\_L6\_2509



N.tet\_L9\_10752



N.tet\_L7\_9045



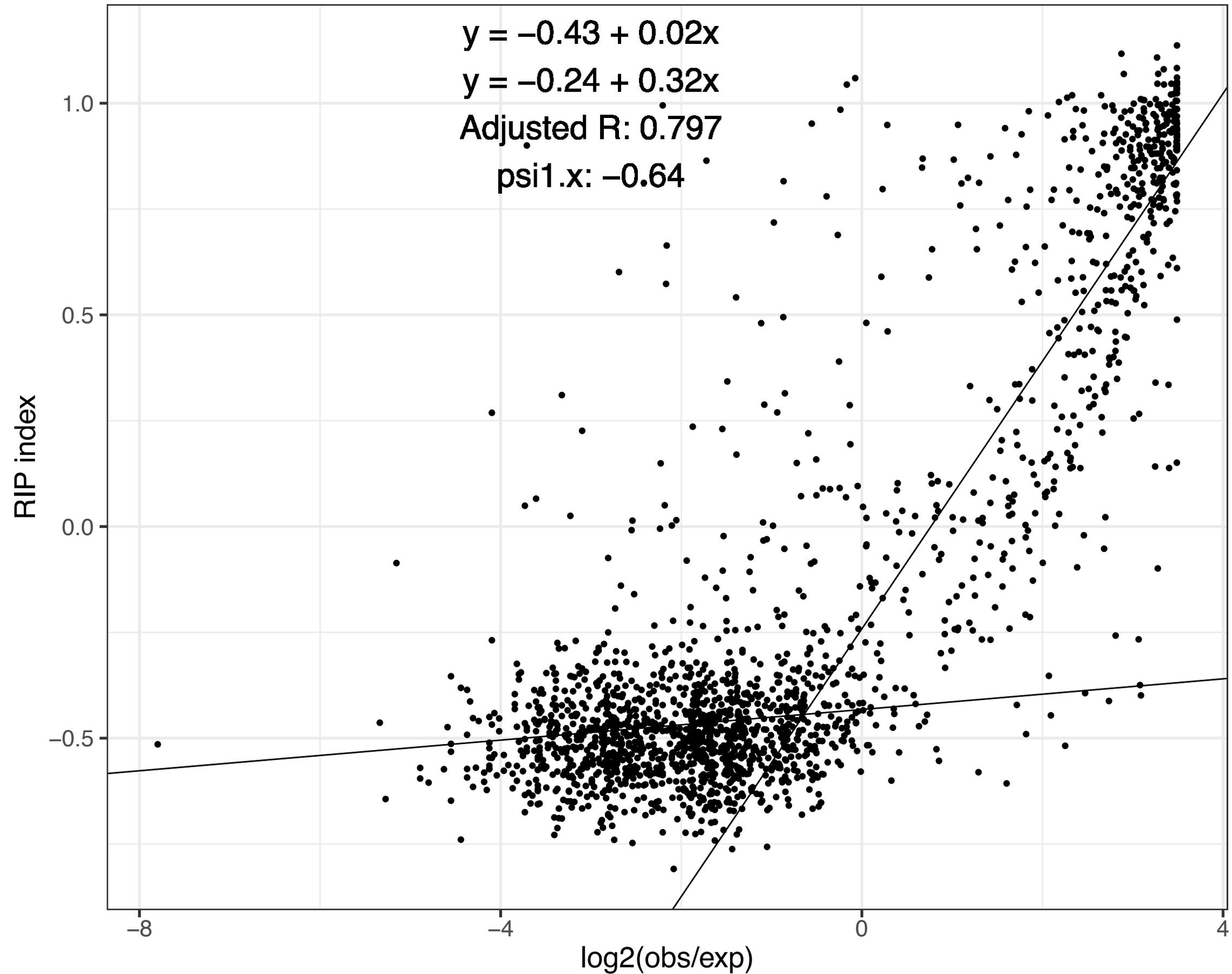
N.tet\_L7\_9046

$$y = -0.43 + 0.02x$$

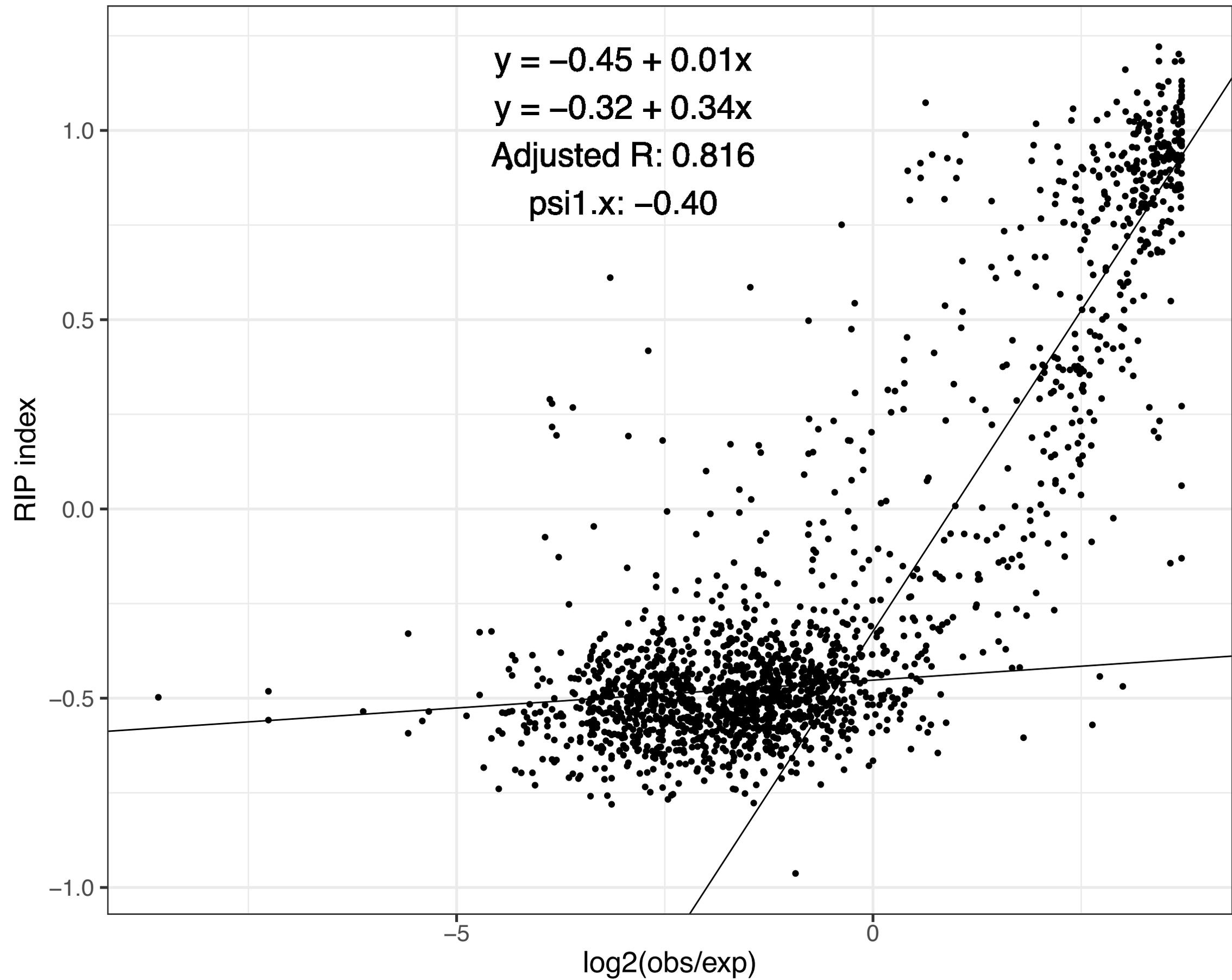
$$y = -0.24 + 0.32x$$

Adjusted R: 0.797

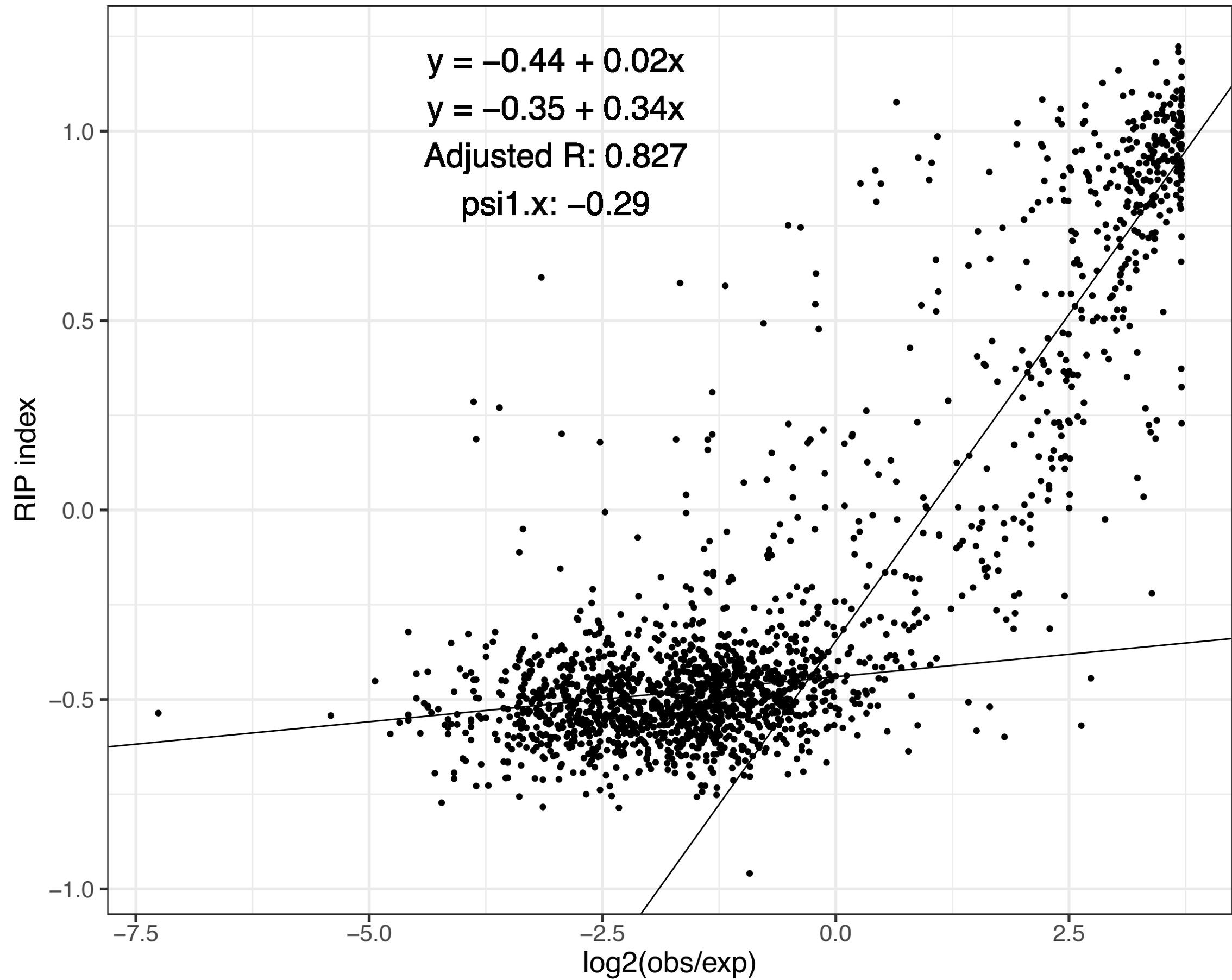
psi1.x: -0.64



N.tet\_L8\_2503

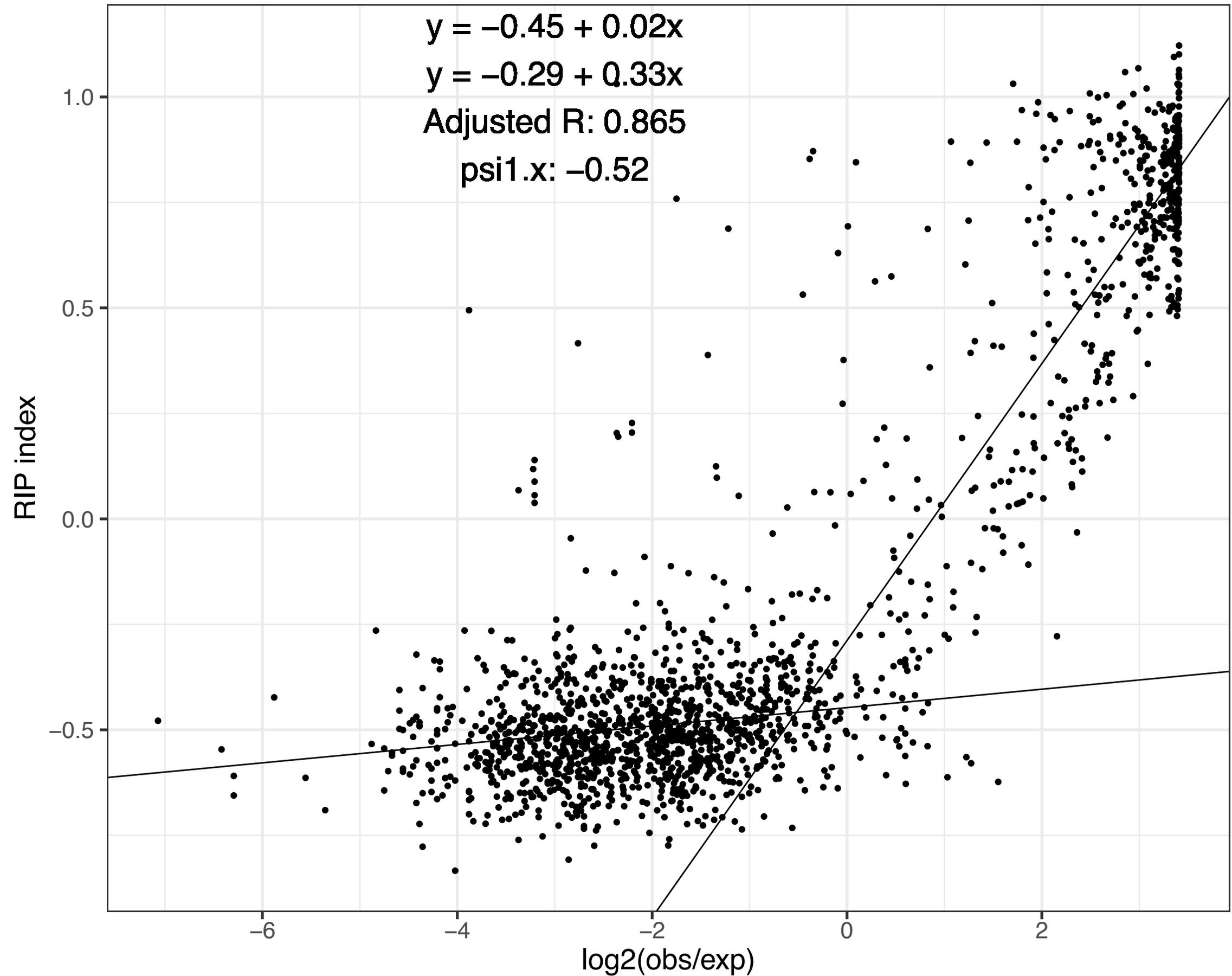


N.tet\_L8\_2504



N.sit\_5940

$y = -0.45 + 0.02x$   
 $y = -0.29 + 0.33x$   
Adjusted R: 0.865  
 $\psi_1.x: -0.52$



N.sit\_5941

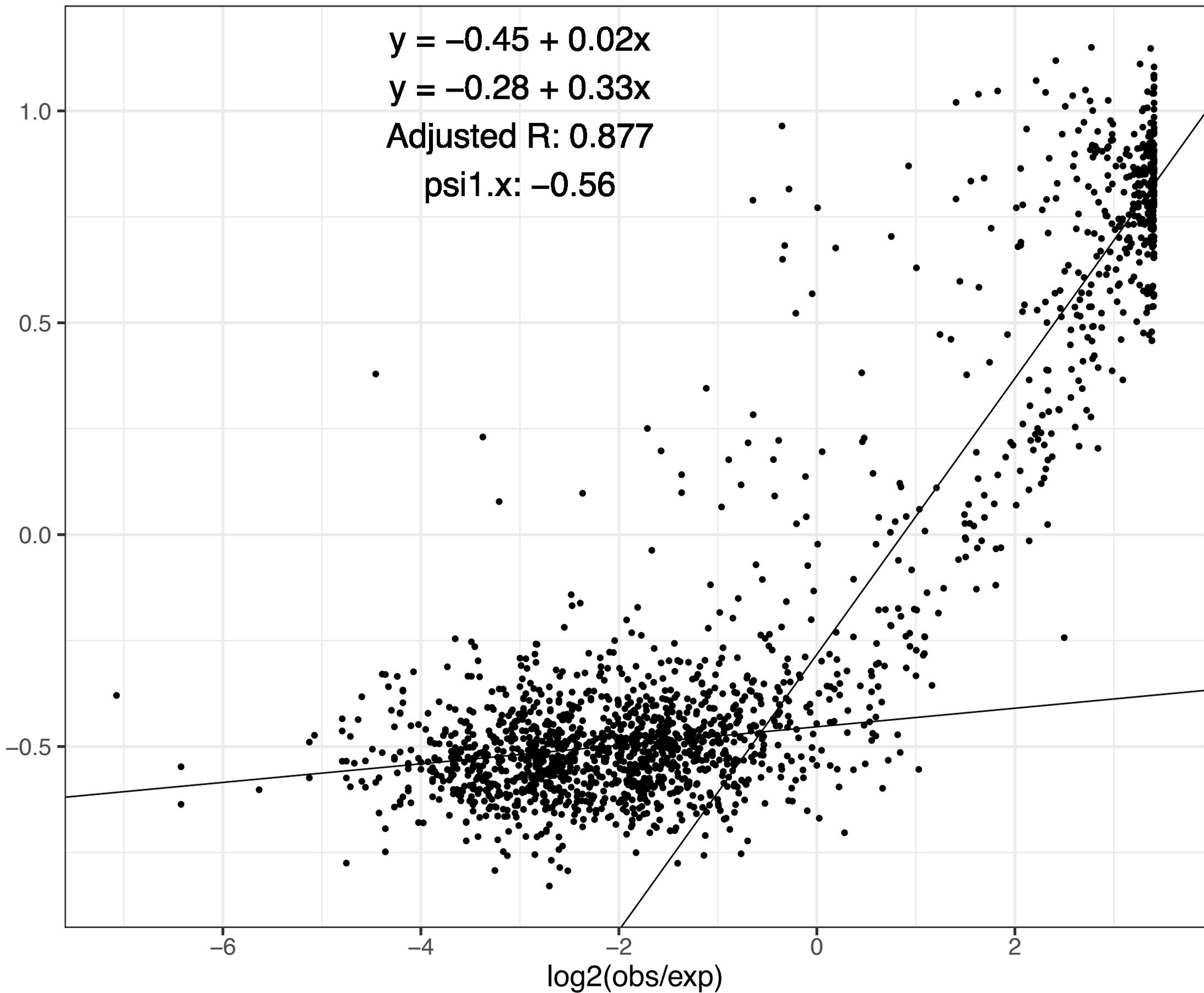
$$y = -0.45 + 0.02x$$

$$y = -0.28 + 0.33x$$

Adjusted R: 0.877

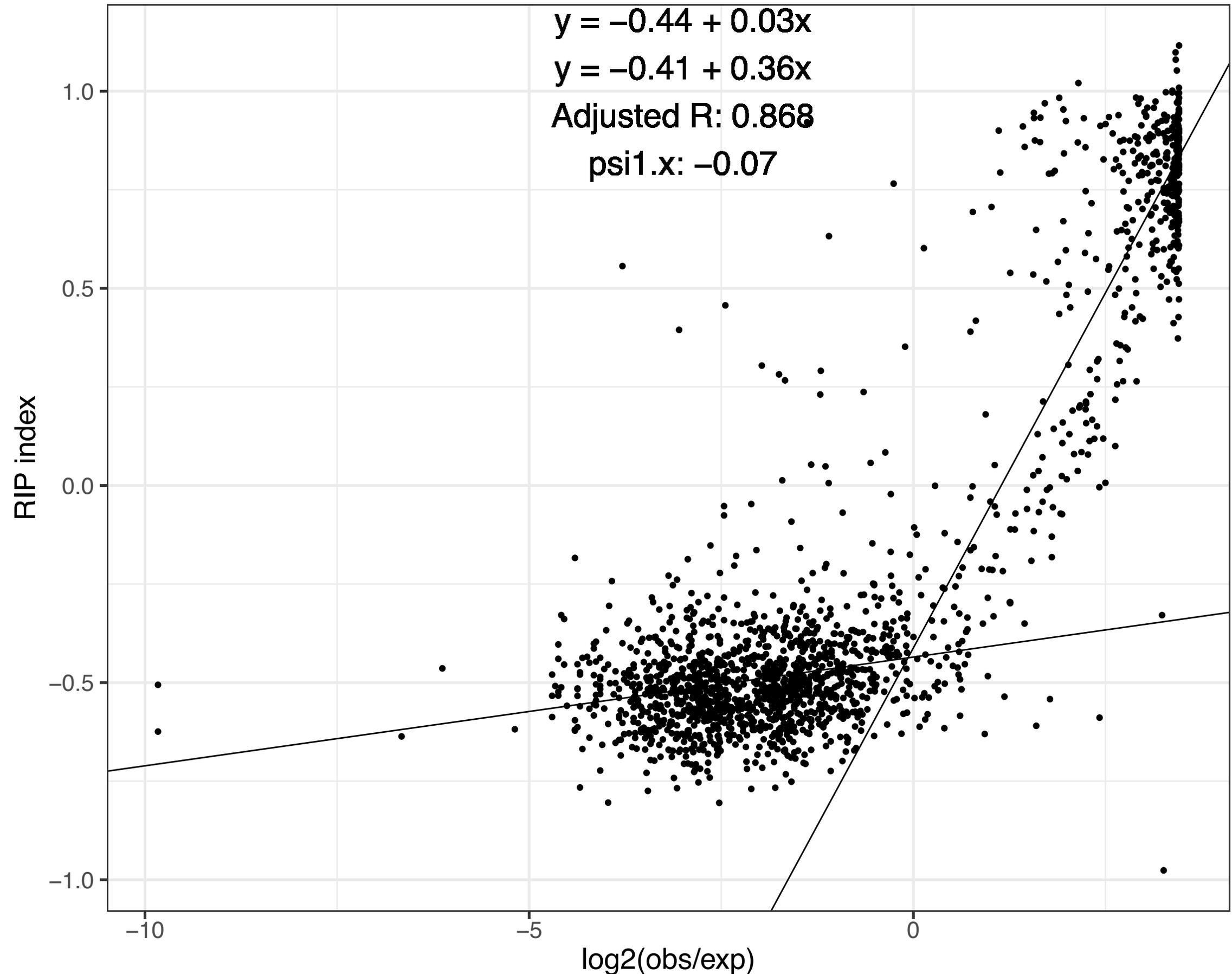
psi1.x: -0.56

RIP index



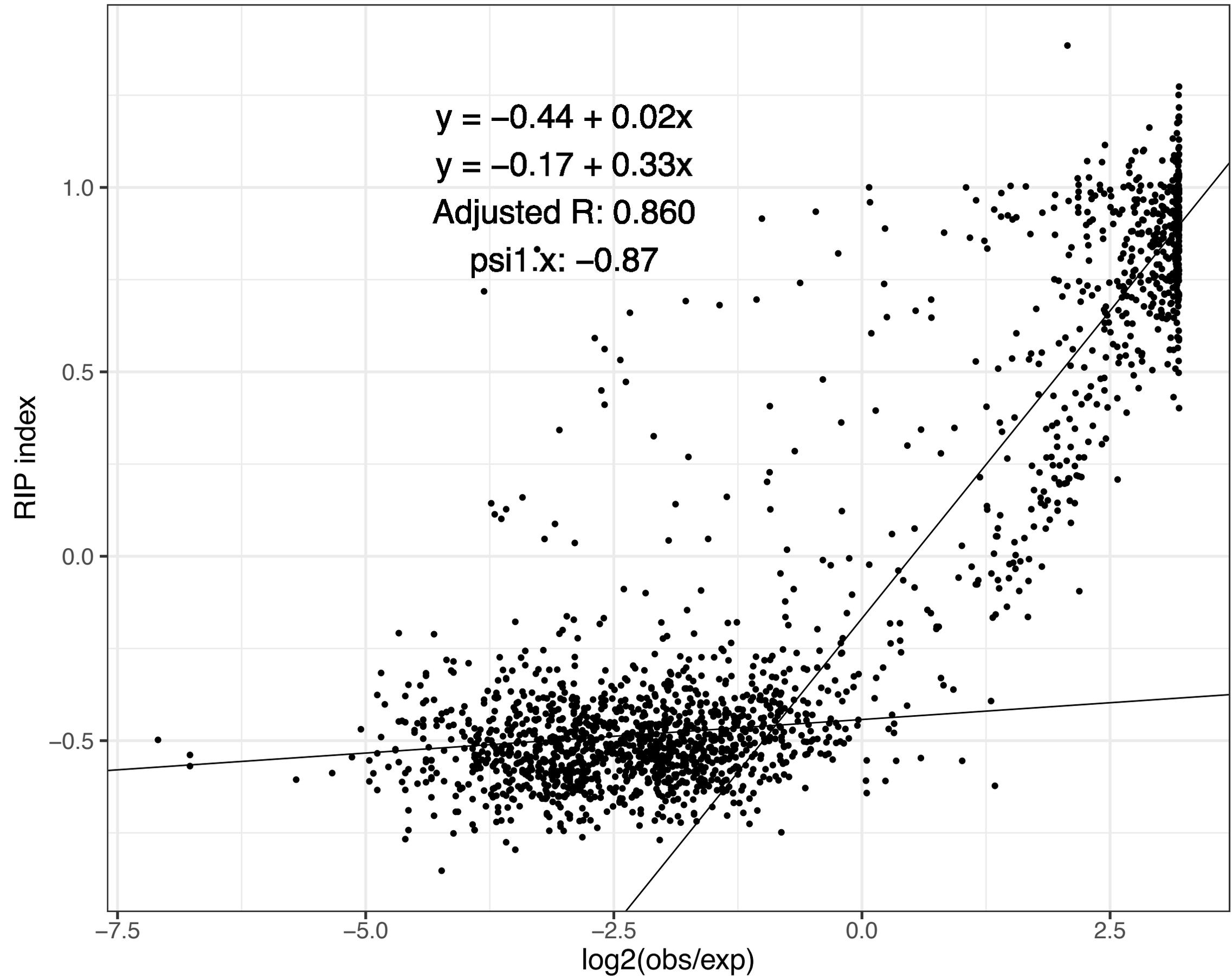
N.sit\_W1426

$y = -0.44 + 0.03x$   
 $y = -0.41 + 0.36x$   
Adjusted R: 0.868  
psi1.x: -0.07

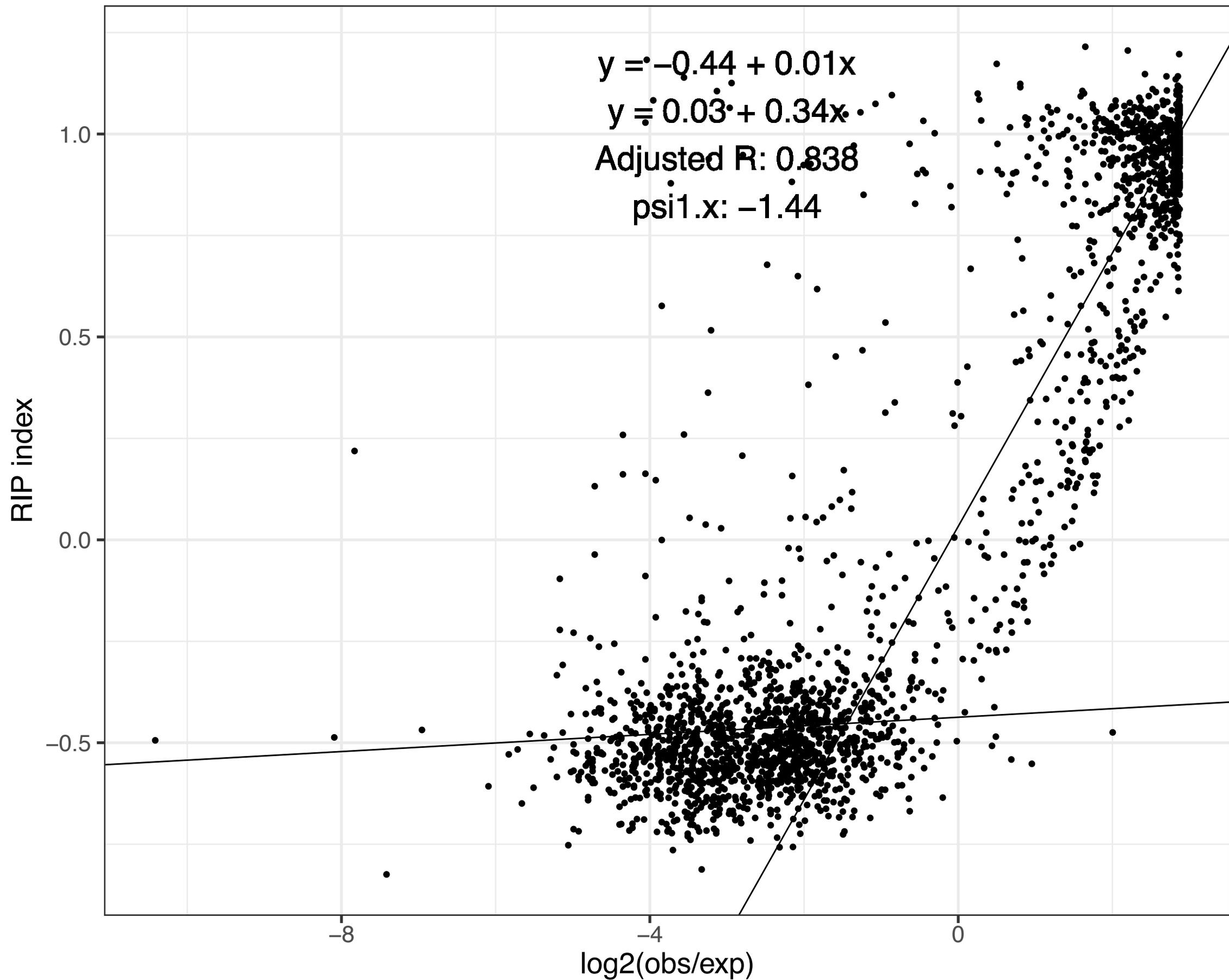


N.sit\_W1434

$y = -0.44 + 0.02x$   
 $y = -0.17 + 0.33x$   
Adjusted R: 0.860  
psi1.x: -0.87

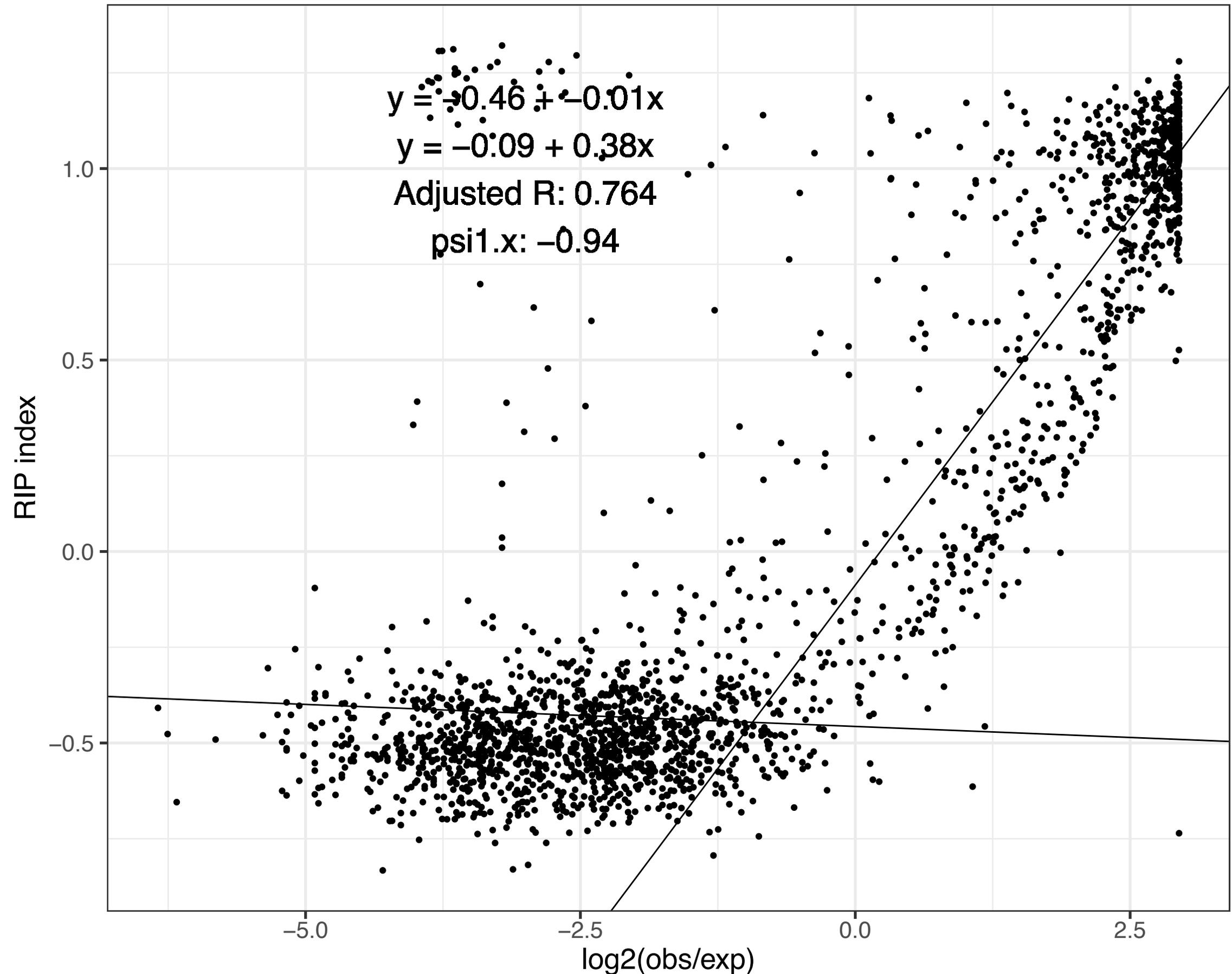


N.his\_8817

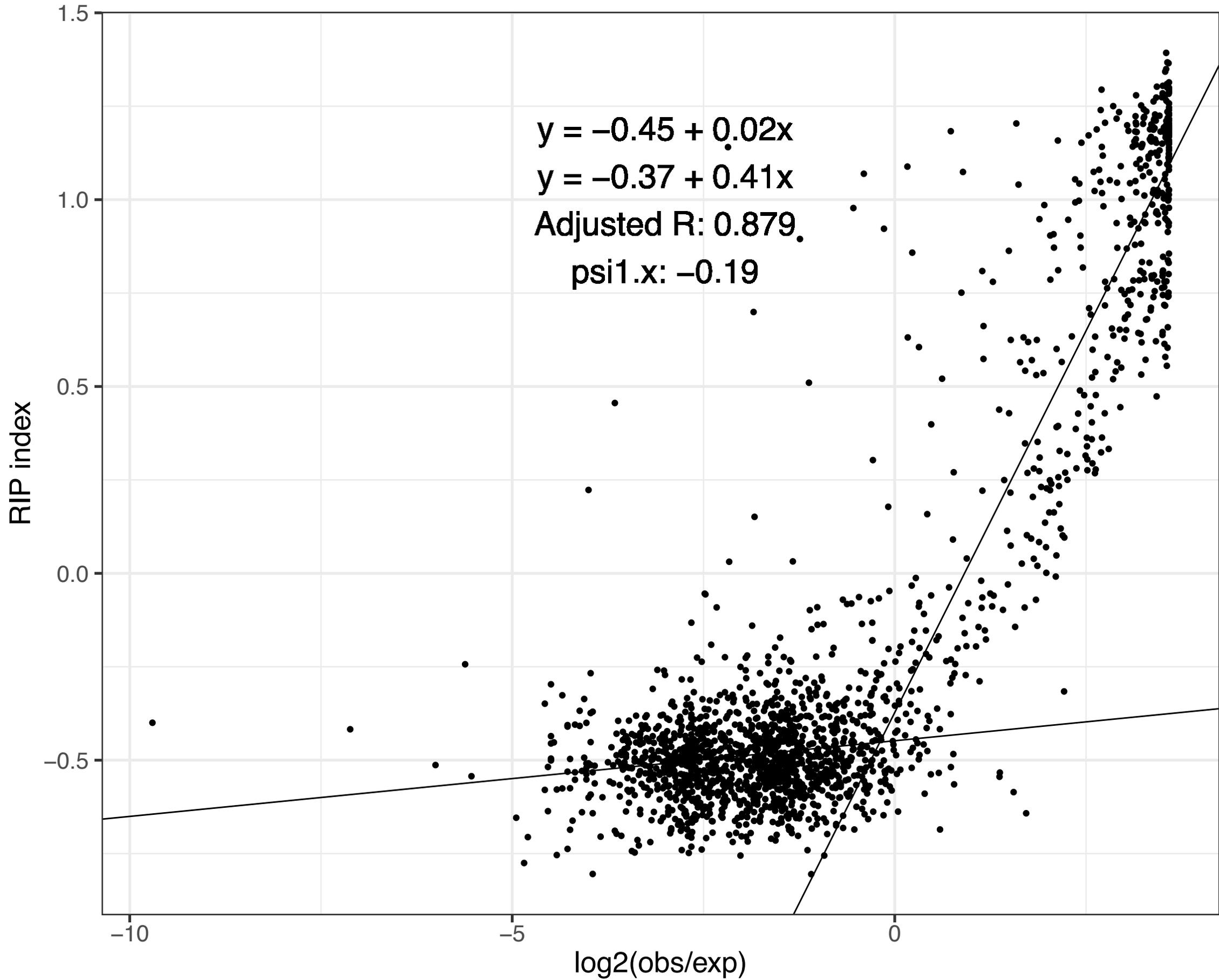


N.cra\_2489

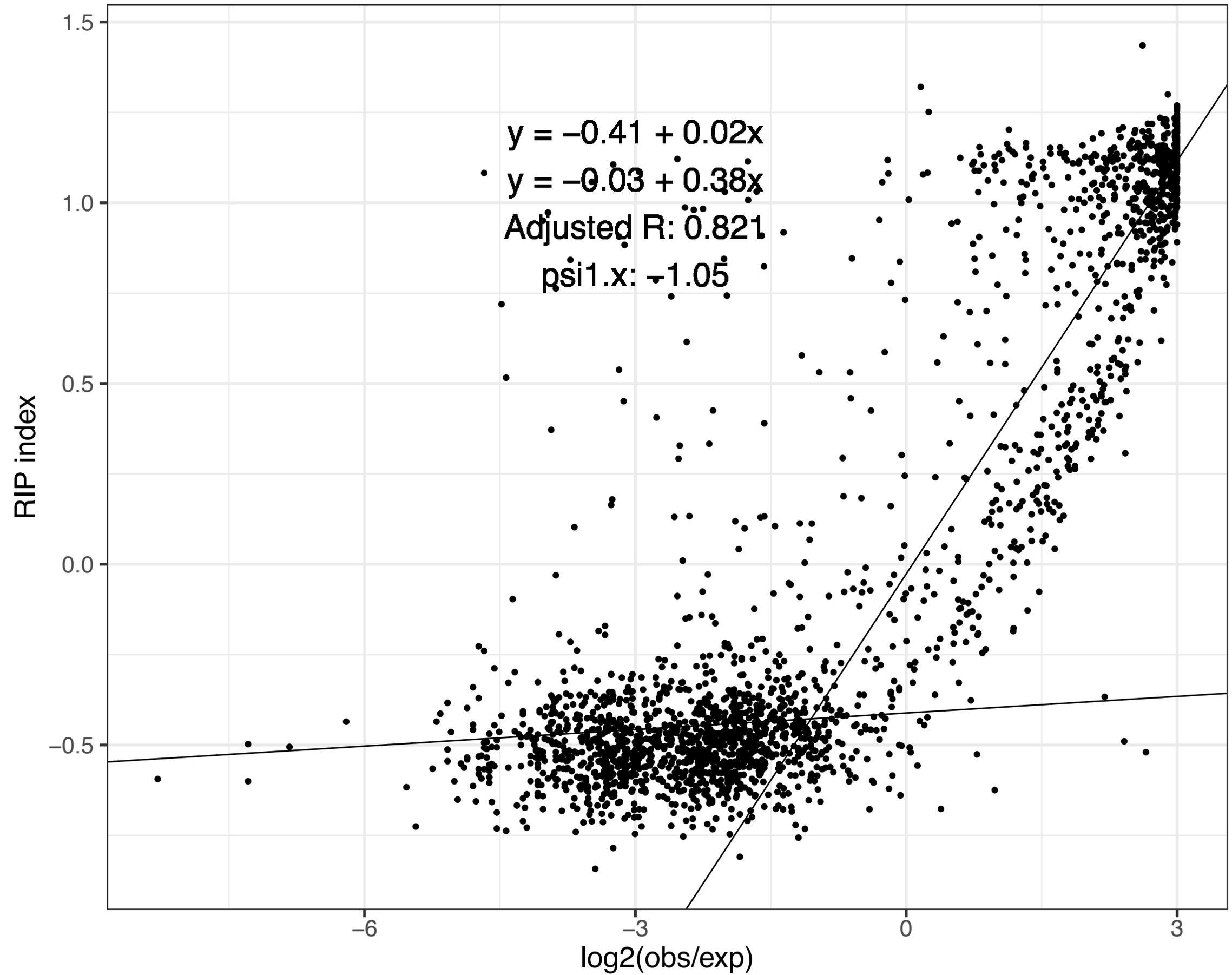
$y = -0.46 + -0.01x$   
 $y = -0.09 + 0.38x$   
Adjusted R: 0.764  
psi1.x: -0.94



N.int\_8767

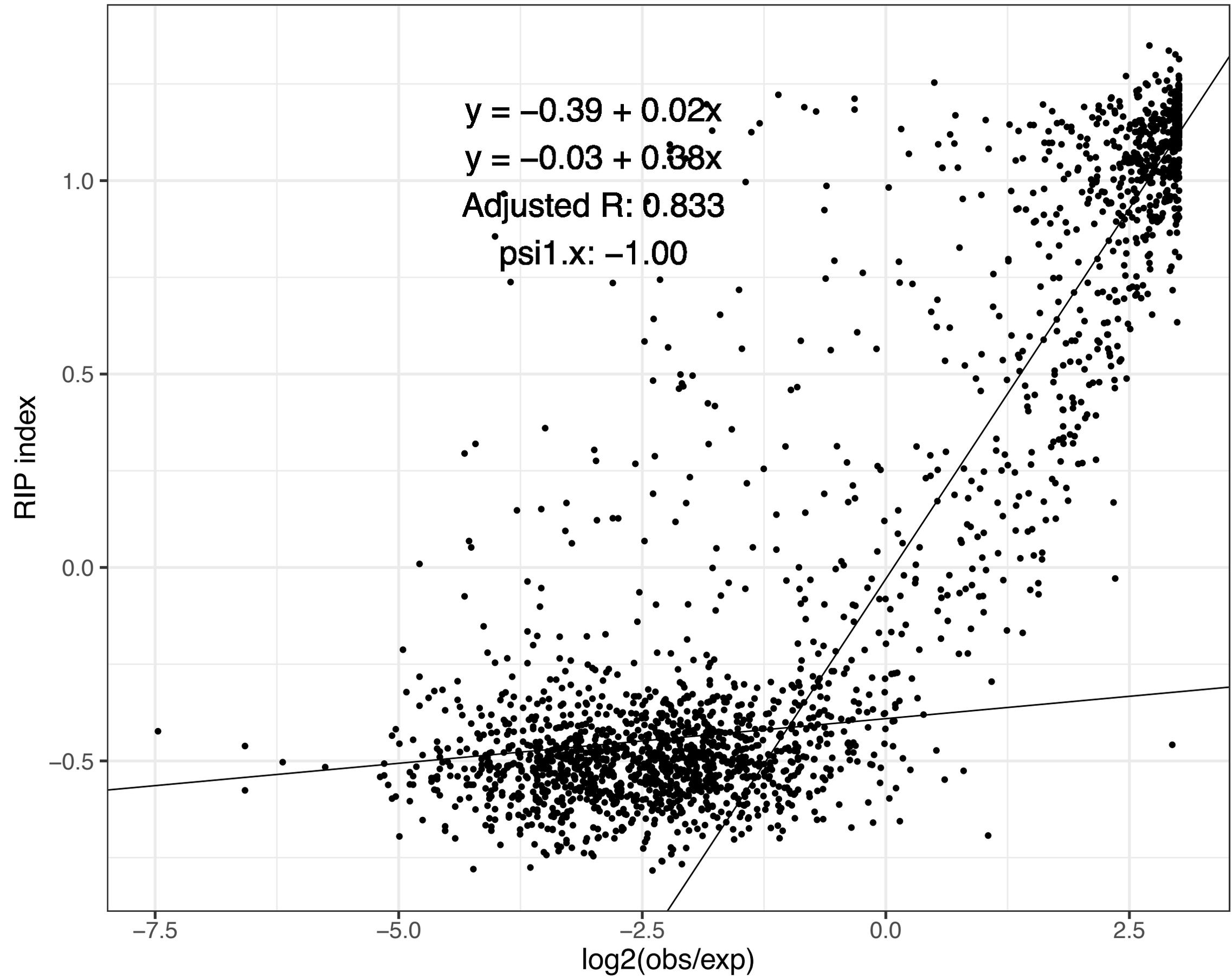


N.int\_8807

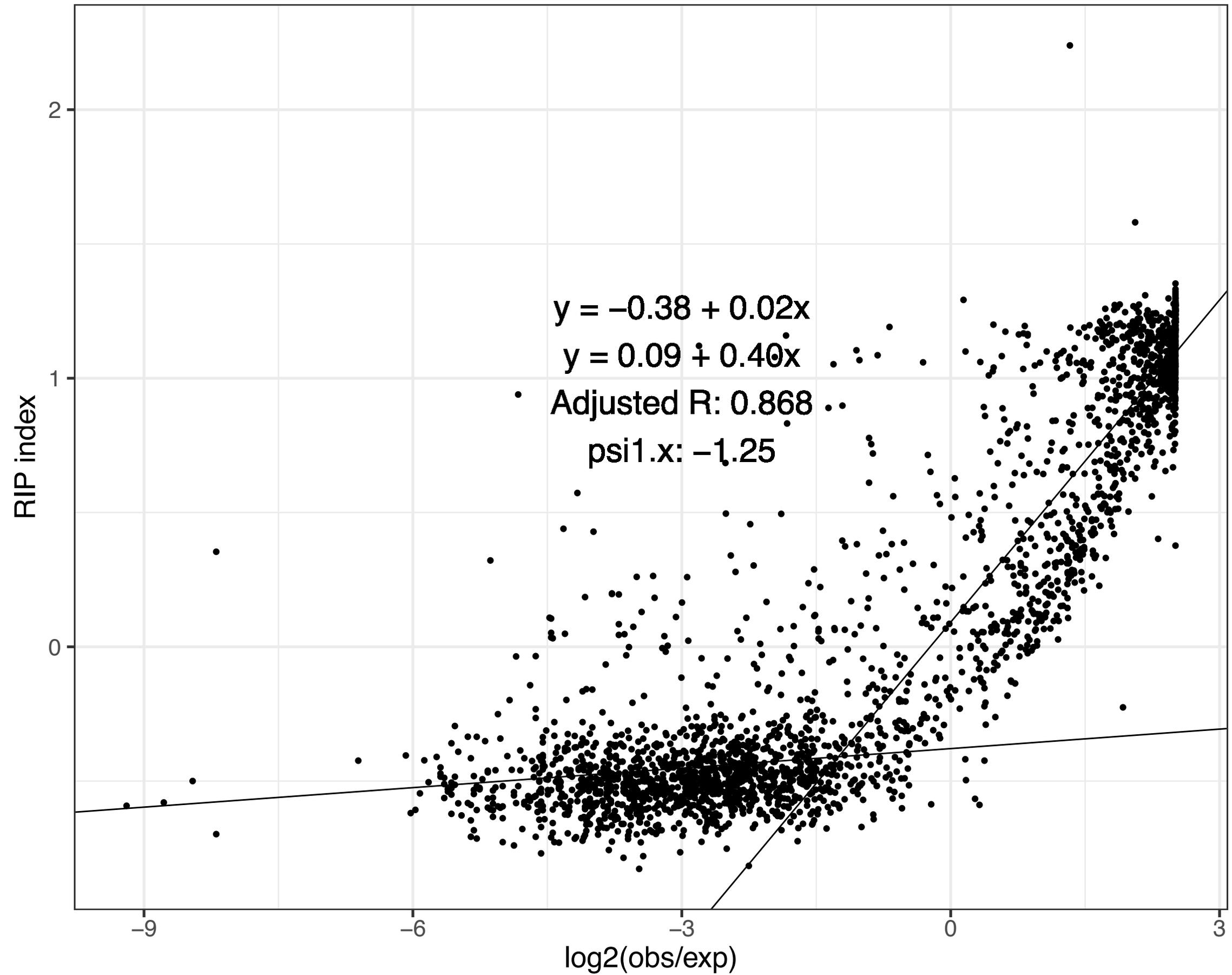


N.int\_8761

$y = -0.39 + 0.02x$   
 $y = -0.03 + 0.38x$   
Adjusted R: 0.833  
psi1.x: -1.00



N.met\_10397



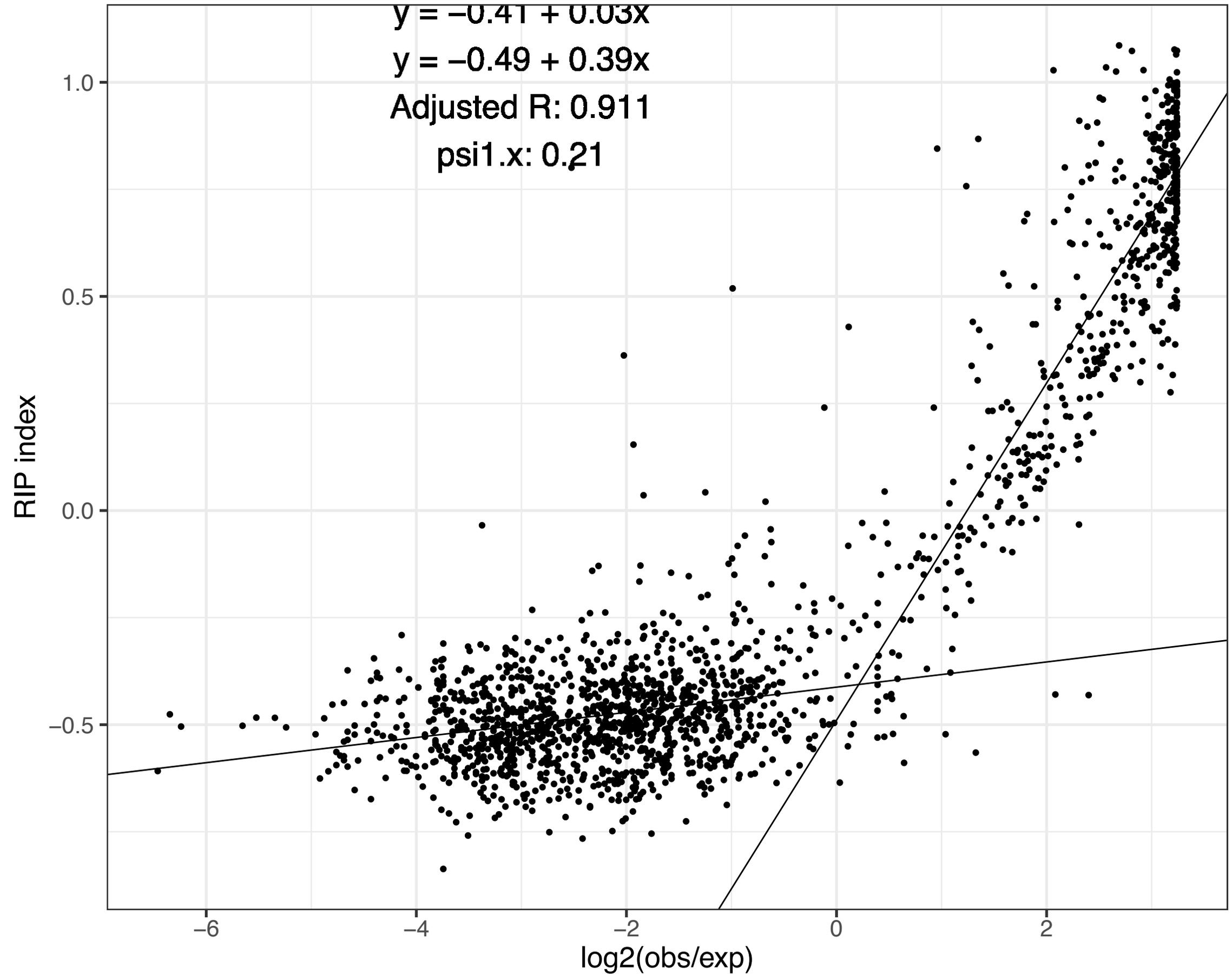
N.dis\_8579

$$y = -0.41 + 0.03x$$

$$y = -0.49 + 0.39x$$

Adjusted R: 0.911

psi1.x: 0.21



**Table S1.** Strain and assembly statistics for 18 *Neurospora* strains and data availability.

Strain/FGSC <sup>a</sup>	Species <sup>b</sup>	Mating type	Number of unitigs	Number of placed unitigs <sup>c</sup>	Assembly size (bp)	N50 contig length (bp)	Mean sequencing coverage	GC content (%)	BUSCO (%) <sup>d</sup>	Genome first referenced in respective publications	Data repository <sup>e</sup>
<b>PacBio</b>											
965a/10752	<i>N. tetrasperma</i> L9	a	24	7	38,575,072	5,668,045	50.1	49.53	99.31	Sun et al (2017)	PRJNA398702
CJ73/9045	<i>N. tetrasperma</i> L7	A	27	7	39,009,030	5,642,919	61.2	49.58	99.31	Sun et al (2017)	PRJNA398702
CJ74/9046	<i>N. tetrasperma</i> L7	a	16	7	38,902,778	5,642,993	55.22	49.54	99.66	Sun et al (2017)	PRJNA398702
CJ85/2503	<i>N. tetrasperma</i> L8	A	29	8	38,526,578	5,531,777	64.6	49.89	99.31	Sun et al (2017)	PRJNA398702
CJ86/2504	<i>N. tetrasperma</i> L8	a	37	8	38,582,854	5,531,742	69.13	49.91	99.31	Sun et al (2017)	PRJNA398702
8817 <sup>f</sup>	<i>N. hispaniola</i>	A	26	7	41,481,169	6,020,640	127.05	47.75	99.31	This study	PRJNA622402
W1426 <sup>g</sup>	<i>N. sitophila</i>	A	20	8	37,879,917	4,685,423	93.47	49.9	99.31	Svedberg et al (2020)	PRJNA649678
5940	<i>N. sitophila</i>	A	8	7	37,949,446	5,431,983	94.19	49.71	99.31	Hosseini et al (2020)	PRJNA505300
5941	<i>N. sitophila</i>	a	25	7	38,368,316	5,449,980	97.5	49.63	99.31	Hosseini et al (2020)	PRJNA505300
W1434 <sup>g</sup>	<i>N. sitophila</i>	A	14	7	39,058,735	5,522,155	80.17	49.33	99.31	Sun et al (2017)	PRJNA398702
8767	<i>N. intermedia</i>	A	35	24	37,915,042	2,232,336	147.51	49.97	98.97	This study	PRJNA622402
8761	<i>N. intermedia</i>	A	27	10	41,248,059	5,337,510	143.34	48.37	99.31	Svedberg et al (2018)	PRJNA486257
8807	<i>N. intermedia</i>	a	14	7	41,318,598	5,974,681	69.28	48.05	99.31	Sun et al (2017), Svedberg et al (2018)	PRJNA398702
10397	<i>N. metzenbergii</i>	a	27	8	43,917,869	5,660,310	86.94	47.75	98.97	This study	PRJNA622402
8579	<i>N. discreta</i>	A	22	7	37,801,234	5,156,816	111.68	49.41	99.31	This study	PRJNA622402
<b>Reference</b>											
2508	<i>N. tetrasperma</i> L6	A	82	na	39,240,259	5,681,182		49.39	99	Ellison et al (2011)	
2509	<i>N. tetrasperma</i> L6	a	308	na	39,220,604	5,677,391		49.65	99.31	Ellison et al (2011)	
2489	<i>N. crassa</i>	A	20	7	41,037,538	6,000,761		48.25	99.66h	Galagan et al (2003)	

a. Strain numbers correspond to the strain ID in the Fungal Genetics Stock Center (FGSC).

b. The four different lineages of *N. tetrasperma* (L6, L9, L7, L8) are herein referred to as different species, according to Corcoran, et al. (2014, 2016) and Menkis, et al. (2009).

c. Number of contigs mapping to the seven chromosomes of the *N. crassa* OR74A reference genome

d. BUSCO (v2/v3) was determined using gVolante with reference gene set Fungi (290 total number of core genes queries against).

e. The completed genome assemblies (including those not previously published; HGAP3\_CRT assembler) are available at <https://doi.org/10.6084/m9.figshare.7346711.v1> (Nguyen, et al., 2022). Raw PacBio reads have been deposited in the Sequence Read Archive as BioProject PRJNA622402.

f. Following the sequencing of strain 8817 (*N. hispaniola*), the sum of contig lengths is 42,505,676 bp. However, we detected that there was about 1 Mbp of bacterial sequence contamination that resulted in 149 contigs. We removed 122 contigs that were of bacterial origin and 1 contig that was the PacBio rolling circle control, leaving 26 contigs that remain in the final assembly.

g. ID numbers from D Jacobson's personal collection (Jacobson et al., 2006).

h. BUSCO for the reference genome for *N. crassa* (GSE71024\_or74a\_12; genome length: 41,037,538 bp in 20 contigs) is 99.7%. N50 is 6,000,761 bp

## References

- Corcoran P., et al. Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus *Neurospora tetrasperma*. *Genome Research* 26: 486-498 (2016).
- Corcoran P., et al. A global multilocus analysis of the model fungus *Neurospora* reveals a single recent origin of a novel genetic system. *Mol Phylogenet Evol* 78: 136-147 (2014).
- Ellison C. E., et al. Massive changes in genome architecture accompany the transition to self-fertility in the filamentous fungus *Neurospora tetrasperma*. *Genetics* 189(1):55-69 (2011).
- Galagan J. E., et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422(6934):859-68 (2003).
- Jacobson, D. J., et al. New findings of *Neurospora* in Europe and comparisons of diversity in temperate climates on continental scales. *Mycologia* 98, 550–559 (2006).
- Menkis A., et al. Phylogenetic and biological species diversity within the *Neurospora tetrasperma* complex. *J Evol Biol* 22: 1923-1936 (2009).
- Nguyen D., et al. Datasets and analysis pipeline deposited for Nguyen et al., (2022) "Transposon- and genome dynamics in the fungal genus *Neurospora*: insights from nearly gapless genome assemblies". Figshare <https://doi.org/10.6084/m9.figshare.c.4310996.v2> (collection of datasets and pipeline, including <https://doi.org/10.6084/m9.figshare.7346711.v1> (assemblies) ).
- Nishimura O., et al. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* 33(22):3635-3637 (2017).
- Sun, Y., et al. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nature Communications* 8(1): 1140 (2017).
- Svedberg, J., et al. Convergent evolution of complex genomic rearrangements in two fungal meiotic drive elements. *Nature Communications* 9(1): 4242 (2018).
- Svedberg, J., et al. An introgressed gene causes meiotic drive in *Neurospora sitophila*. *Proceedings of the National Academy of Sciences* 118 (2021).

**Table S2.** Genomic transposable element content for 18 *Neurospora* strains.

Species name	Genome identifier	Genome length (bp)	Interspersed Repeats length (bp)	Interspersed Repeats (%)	Percent of genome							Length of sequences						
					LTR	LINE	SINE	DNA	Unclassified	Simple Repeats	Low Complexity	LTR	LINE	SINE	DNA	Unclassified	Simple Repeats	Low Complexity
<i>N. tetrasperma</i> L6	2508	39240259	3516218	8.96	5.17	1.04	0.06	0.41	2.28	1.63	0.26	2027686	407065	23797	162443	895227	638720	101802
<i>N. tetrasperma</i> L6	2509	39220604	3196702	8.15	4.56	1.01	0.07	0.37	2.15	1.56	0.25	1786706	395781	25628	144879	843708	612684	98039
<i>N. tetrasperma</i> L9	10752pb	38575072	3476437	9.01	5.27	1.05	0.07	0.37	2.26	1.64	0.27	2032992	403992	25144	142676	871633	630744	102591
<i>N. tetrasperma</i> L7	9045pb	39009030	3393923	8.7	5.18	0.86	0.07	0.39	2.2	1.63	0.26	2021269	336633	25599	150593	859829	635864	102278
<i>N. tetrasperma</i> L7	9046pb	38902778	3452713	8.88	5.03	1.18	0.07	0.37	2.23	1.64	0.26	1958563	458549	25707	143025	866869	637179	102614
<i>N. tetrasperma</i> L8	2503pb	38526578	2938288	7.63	3.82	1.06	0.07	0.33	2.35	1.66	0.27	1471521	408086	25287	127498	905896	638931	103115
<i>N. tetrasperma</i> L8	2504pb	38582854	2924517	7.58	3.74	1.18	0.07	0.41	2.19	1.66	0.26	1442393	454632	25388	157693	844411	639906	100371
<i>N. hispaniola</i>	8817pb	41481169	5618515	13.54	8.68	1.5	0.06	0.75	2.56	1.77	0.26	3601704	621440	25539	309841	1059991	733141	108558
<i>N. sitophila</i>	5940pb	37949446	3578890	9.43	4.96	1.87	0.07	0.69	1.84	1.76	0.26	1884175	709294	25109	260884	699428	667099	97090
<i>N. sitophila</i>	5941pb	38368316	3578710	9.33	4.82	1.91	0.07	0.71	1.82	1.76	0.25	1850457	733884	25053	271302	698014	675999	97529
<i>N. sitophila</i>	W1426pb	37879917	3452152	9.11	4.34	2.04	0.07	0.51	2.16	1.72	0.25	1643845	772772	25033	192900	817602	650821	96186
<i>N. sitophila</i>	W1434pb	39058735	4254038	10.89	6.1	1.91	0.06	0.76	2.07	1.73	0.25	2381859	744544	25343	295466	806826	676423	99088
<i>N. crassa</i>	GSE71024	41037538	5342783	13.02	7.11	1.83	0.06	1.1	2.91	1.85	0.28	2918994	750422	25746	452809	1194812	758159	114473
<i>N. intermedia</i>	8767pb	37915042	3173439	8.37	4.87	1.11	0.07	0.24	2.08	1.77	0.27	1847606	421962	25546	90571	787754	669584	100533
<i>N. intermedia</i>	8807pb	41318598	5154185	12.47	7.11	1.85	0.06	0.79	2.66	1.83	0.28	2938362	765932	25424	325566	1098901	754422	113662
<i>N. intermedia</i>	8761pb	41248059	5101337	12.37	7.73	1.4	0.06	0.65	2.53	1.75	0.28	3186734	577175	25559	266317	1045552	722814	113612
<i>N. metzenbergii</i>	10397pb	43917869	7656512	17.43	8.77	2.74	0.06	1.17	4.69	1.61	0.25	3853784	1204052	25872	514795	2058009	705118	111472
<i>N. discreta</i>	8579pb	37801234	3954853	10.46	6.55	0.94	0.07	0.46	2.45	1.6	0.25	2476627	355849	24975	172761	924641	603076	93349

**Table S3.** Counts of *Neurospora* pairwise lineage-specific transposable elements. The table is read as: genome of Species 1 has X lineage-specific insertions with respect to the genome of Species 2.

Genome		Species 2																	
		N.tet L6 2508	N.tet L6 2509	N.tet L9 10752	N.tet L7 9045	N.tet L7 9046	N.tet L8 2503	N.tet L8 2504	N.his 8817	N.sit 5940	N.sit 5941	N.sit W1426	N.sit W1434	N.cra 2489	N.int 8767	N.int 8807	N.int 8761	N.met 10397	N.dis 8579
Species 1	N.tet L6 2508	NA	5	2	4	4	3	4	1	2	4	2	3	2	1	1	2	1	3
	N.tet L6 2509	8	NA	6	8	7	9	6	1	2	3	0	1	1	0	0	3	2	0
	N.tet L9 10752	4	5	NA	3	1	5	3	5	1	2	1	3	3	1	2	0	3	0
	N.tet L7 9045	4	3	7	NA	10	7	6	3	2	1	1	0	1	3	4	0	5	0
	N.tet L7 9046	2	2	4	9	NA	7	3	1	3	2	3	2	1	3	2	3	4	1
	N.tet L8 2503	4	5	2	5	4	NA	9	4	1	1	0	1	1	2	1	1	3	0
	N.tet L8 2504	5	5	2	4	3	9	NA	3	2	2	0	1	0	0	0	0	1	1
	N.his 8817	1	3	3	3	6	2	2	NA	1	2	3	2	4	4	6	3	3	1
	N.sit 5940	2	1	2	2	2	0	2	2	NA	13	8	11	0	0	1	1	2	0
	N.sit 5941	0	0	0	1	1	2	2	0	17	NA	12	15	0	1	0	1	2	0
	N.sit W1426	3	3	4	2	3	4	2	5	17	16	NA	23	2	5	4	4	1	1
	N.sit W1434	1	4	5	7	6	1	2	2	14	17	13	NA	3	7	4	5	3	1
	N.cra 2489	8	5	7	5	4	2	3	6	4	1	4	4	NA	12	10	9	6	3
	N.int 8767	1	2	3	2	1	1	1	2	0	1	2	1	3	NA	3	2	0	0
	N.int 8807	5	3	3	4	4	4	6	2	4	4	4	6	7	5	NA	2	5	1
	N.int 8761	10	5	8	5	7	13	11	5	5	5	6	3	8	3	10	NA	3	0
	N.met 10397	14	3	6	11	9	4	3	5	9	9	3	1	6	8	5	9	NA	5
N.dis 8579	0	0	1	0	0	1	1	1	1	1	1	1	2	1	1	1	2	NA	



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	<b>Table S4.</b> Number of insertions by transposable element subfamily in each pairwise comparison in <i>Neurospora</i> .														
2	Comparison	ncra_LTR_104	ncra_LTR_69	Tad1-1	ncra_Tad1_01	ntet_LTR_45	ncra_LTR_53	ntet_Tad1_01	ncra_Tad1_06	ntet_LTR_28	ncra_LTR_63	Tad3-2	ncra_Gypsy_10	ntet_LTR_35	ncra_LTR_05
37	2503_10397	0	0	0	2	0	0	0	0	0	0	0	0	0	0
38	2503_10752	0	0	0	1	0	0	0	0	0	0	0	0	0	0
39	2503_2504	0	0	0	4	0	0	0	0	0	0	0	0	0	0
40	2503_2508	0	0	0	2	0	0	1	0	0	0	0	0	0	0
41	2503_2509	0	0	0	2	0	0	1	0	0	0	0	0	0	0
42	2503_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	2503_8761	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	2503_8767	0	0	0	1	0	0	1	0	0	0	0	0	0	0
45	2503_8807	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	2503_8817	0	0	0	1	0	0	1	0	0	0	0	0	0	0
47	2503_9045	0	0	0	2	0	0	0	0	0	0	0	0	0	0
48	2503_9046	0	0	0	2	0	0	0	0	0	0	0	0	0	0
49	2503_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50	2503_5941	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	2503_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52	2503_W1426	0	0	0	0	0	0	0	0	0	0	0	0	0	0
53	2503_W1434	0	0	0	0	0	0	1	0	0	0	0	0	0	0
54	2504_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
55	2504_10752	0	0	0	1	0	0	0	0	0	0	0	0	0	0
56	2504_2503	0	0	0	4	0	0	0	0	0	0	0	0	0	0
57	2504_2508	0	0	0	2	0	0	1	0	0	0	0	0	0	0
58	2504_2509	0	0	0	2	0	0	1	0	0	0	0	0	0	0
59	2504_8579	0	0	0	0	0	0	1	0	0	0	0	0	0	0
60	2504_8761	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61	2504_8767	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	2504_8807	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63	2504_8817	0	0	0	1	0	0	2	0	0	0	0	0	0	0
64	2504_9045	0	0	0	2	0	0	0	0	0	0	0	0	0	0
65	2504_9046	0	0	0	2	0	0	0	0	0	0	0	0	0	0
66	2504_5940	0	0	0	1	0	0	0	0	0	0	0	0	0	0
67	2504_5941	0	0	0	1	0	0	0	0	0	0	0	0	0	0
68	2504_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
69	2504_W1426	0	0	0	0	0	0	0	0	0	0	0	0	0	0
70	2504_W1434	0	0	0	0	0	0	1	0	0	0	0	0	0	0









	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	<b>Table S4.</b> Number of insertions by transposable element subfamily in each pairwise comparison in <i>Neurospora</i> .														
2	Comparison	ncra_LTR_104	ncra_LTR_69	Tad1-1	ncra_Tad1_01	ntet_LTR_45	ncra_LTR_53	ntet_Tad1_01	ncra_Tad1_06	ntet_LTR_28	ncra_LTR_63	Tad3-2	ncra_Gypsy_10	ntet_LTR_35	ncra_LTR_05
216	9046_10397	0	1	0	1	0	0	0	0	0	0	0	0	0	0
217	9046_10752	0	0	0	2	0	0	0	0	0	0	0	0	0	0
218	9046_2503	0	0	0	3	1	0	0	0	0	0	0	0	0	0
219	9046_2504	0	0	0	2	1	0	0	0	0	0	0	0	0	0
220	9046_2508	0	0	0	1	1	0	0	0	0	0	0	0	0	0
221	9046_2509	0	0	0	1	1	0	0	0	0	0	0	0	0	0
222	9046_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
223	9046_8761	0	0	0	2	0	0	1	0	0	0	0	0	0	0
224	9046_8767	0	0	0	1	0	0	1	0	0	0	0	0	0	0
225	9046_8807	0	0	1	0	0	0	0	0	0	0	0	0	0	0
226	9046_8817	0	0	0	0	0	0	1	0	0	0	0	0	0	0
227	9046_9045	0	0	0	5	0	0	0	0	0	0	0	0	0	0
228	9046_5940	0	0	0	1	0	0	0	0	0	0	0	0	0	0
229	9046_5941	0	0	0	1	0	0	0	0	0	0	0	0	0	0
230	9046_2489	0	0	0	0	0	0	1	0	0	0	0	0	0	0
231	9046_W1426	0	0	0	2	1	0	0	0	0	0	0	0	0	0
232	9046_W1434	0	1	0	1	0	0	0	0	0	0	0	0	0	0
233	5940_10397	0	0	0	0	0	0	0	0	0	1	0	0	0	0
234	5940_10752	0	0	0	0	0	0	0	0	0	0	0	0	0	0
235	5940_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	0
236	5940_2504	0	0	0	1	0	0	0	0	0	0	0	0	0	0
237	5940_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	0
238	5940_2509	0	0	0	0	0	0	0	0	0	0	0	0	0	0
239	5940_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
240	5940_8761	0	0	0	0	0	0	0	0	0	0	0	0	0	0
241	5940_8767	0	0	0	0	0	0	0	0	0	0	0	0	0	0
242	5940_8807	0	0	0	0	0	0	0	0	0	0	0	0	0	0
243	5940_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
244	5940_9045	0	1	0	1	0	0	0	0	0	0	0	0	0	0
245	5940_9046	0	1	0	0	0	0	0	0	0	0	0	0	0	0
246	5940_5941	0	0	1	6	0	0	0	0	0	0	1	0	0	0
247	5940_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
248	5940_W1426	0	0	1	4	0	0	0	0	0	0	1	0	0	0
249	5940_W1434	0	0	1	6	0	0	0	1	0	0	1	0	0	0



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	<b>Table S4.</b> Number of insertions by transposable element subfamily in each pairwise comparison in <i>Neurospora</i> .														
2	Comparison	ncra_LTR_104	ncra_LTR_69	Tad1-1	ncra_Tad1_01	ntet_LTR_45	ncra_LTR_53	ntet_Tad1_01	ncra_Tad1_06	ntet_LTR_28	ncra_LTR_63	Tad3-2	ncra_Gypsy_10	ntet_LTR_35	ncra_LTR_05
285	W1426_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
286	W1426_10752	0	0	0	1	0	0	0	0	0	0	0	0	0	0
287	W1426_2503	0	0	0	0	0	0	0	1	0	0	0	0	0	0
288	W1426_2504	0	0	0	0	0	0	0	1	0	0	0	0	0	0
289	W1426_2508	0	0	0	1	0	0	0	0	0	0	1	0	0	0
290	W1426_2509	0	0	1	1	0	0	0	0	0	0	0	0	0	0
291	W1426_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
292	W1426_8761	0	0	1	1	0	0	0	0	0	0	1	0	0	0
293	W1426_8767	0	0	0	2	0	0	0	0	0	0	1	0	0	0
294	W1426_8807	0	0	0	0	0	0	0	0	0	0	1	0	0	0
295	W1426_8817	0	0	1	2	0	0	0	1	0	0	0	0	0	0
296	W1426_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
297	W1426_9046	0	0	0	0	0	0	0	0	0	0	1	0	0	0
298	W1426_5940	0	0	1	8	1	0	0	1	0	0	1	0	0	0
299	W1426_5941	0	0	1	8	0	0	0	1	0	0	1	0	0	0
300	W1426_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
301	W1426_W1434	0	0	2	10	0	0	0	2	0	0	1	0	0	0
302	W1434_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
303	W1434_10752	0	0	0	1	0	0	0	1	0	0	0	0	0	0
304	W1434_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	0
305	W1434_2504	0	0	0	0	0	0	0	0	0	0	0	0	0	0
306	W1434_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	0
307	W1434_2509	0	0	1	0	0	0	0	1	0	0	0	0	0	0
308	W1434_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
309	W1434_8761	0	0	1	0	0	0	0	1	0	0	1	0	0	0
310	W1434_8767	1	0	0	1	0	0	0	0	0	0	0	0	0	0
311	W1434_8807	0	0	1	1	0	0	0	0	0	0	1	0	0	0
312	W1434_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
313	W1434_9045	0	0	0	1	0	0	0	0	0	0	0	0	0	0
314	W1434_9046	0	0	0	1	0	0	0	1	0	0	0	0	0	0
315	W1434_5940	0	0	1	7	0	0	0	1	0	0	1	0	0	0
316	W1434_5941	0	0	1	7	0	0	0	1	0	0	1	0	0	0
317	W1434_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
318	W1434_W1426	1	0	1	7	0	0	0	0	0	0	1	0	0	0
319	Total	26	45	40	353	20	19	43	33	4	7	22	1	9	12

	A	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	Table S4. Number of ir													
2	Comparison	ntet_Gypsy_01	ntet_LTR_14	ntet_LTR_18	ncra_LTR_04	ncra_LTR_75	ntet_LTR_37	ncra_LTR_116	ncra_LTR_101	ncra_LTR_64	ncra_Gypsy_01	ncra_Gypsy_08	ntet_LTR_27	ntet_Tad1_02
3	10397_10752	0	0	0	0	0	0	0	0	0	0	0	0	0
4	10397_2503	0	0	0	0	0	0	0	0	0	0	0	0	0
5	10397_2504	0	0	0	0	0	0	0	0	0	0	0	0	0
6	10397_2508	0	0	0	0	0	0	0	0	0	0	0	0	0
7	10397_2509	0	0	0	0	0	0	0	0	0	0	0	0	0
8	10397_8579	0	0	0	0	0	0	0	0	0	0	0	0	0
9	10397_8761	0	0	0	0	0	0	0	0	0	0	0	0	0
10	10397_8767	0	0	0	0	0	0	0	0	0	0	0	0	0
11	10397_8807	0	0	0	0	0	0	0	0	0	0	0	0	0
12	10397_8817	0	0	0	0	0	0	0	0	0	0	0	0	0
13	10397_9045	0	0	0	0	0	0	0	0	0	0	0	0	0
14	10397_9046	0	0	0	0	0	0	0	0	0	0	0	0	0
15	10397_5940	0	0	0	0	0	0	0	0	0	0	0	0	0
16	10397_5941	0	0	0	0	0	0	0	0	0	0	0	0	0
17	10397_2489	0	0	0	0	0	0	0	0	0	0	0	0	0
18	10397_W1426	0	0	0	0	0	0	0	0	0	0	0	0	0
19	10397_W1434	0	0	0	0	0	0	0	0	0	0	0	0	0
20	10752_10397	0	0	0	0	0	0	0	0	0	0	0	0	0
21	10752_2503	1	1	1	0	0	0	0	0	0	0	0	0	0
22	10752_2504	1	0	0	0	0	0	0	0	0	0	0	0	0
23	10752_2508	1	0	0	1	0	0	0	0	0	0	0	0	0
24	10752_2509	1	0	0	0	1	1	0	0	0	0	0	0	0
25	10752_8579	0	0	0	0	0	0	0	0	0	0	0	0	0
26	10752_8761	0	0	0	0	0	0	0	0	0	0	0	0	0
27	10752_8767	0	0	0	0	0	0	0	0	0	0	0	0	0
28	10752_8807	0	1	1	0	0	0	0	0	0	0	0	0	0
29	10752_8817	0	0	1	0	0	2	0	0	0	0	0	0	0
30	10752_9045	1	0	1	0	0	0	0	0	0	0	0	0	0
31	10752_9046	1	0	0	0	0	0	0	0	0	0	0	0	0
32	10752_5940	0	0	1	0	0	0	0	0	0	0	0	0	0
33	10752_5941	0	0	1	0	0	0	0	0	0	0	0	0	0
34	10752_2489	0	1	0	0	0	0	0	0	0	0	0	0	0
35	10752_W1426	0	0	1	0	0	0	0	0	0	0	0	0	0
36	10752_W1434	0	0	1	0	0	0	1	0	0	0	0	0	0















	A	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	Table S4. Number of ir													
2	Comparison	ntet_Gypsy_01	ntet_LTR_14	ntet_LTR_18	ncra_LTR_04	ncra_LTR_75	ntet_LTR_37	ncra_LTR_116	ncra_LTR_101	ncra_LTR_64	ncra_Gypsy_01	ncra_Gypsy_08	ntet_LTR_27	ntet_Tad1_02
285	W1426_10397	0	0	0	0	0	0	0	0	0	0	0	0	0
286	W1426_10752	0	0	0	0	0	0	0	0	0	1	0	0	0
287	W1426_2503	0	0	0	0	0	0	0	0	0	1	0	0	0
288	W1426_2504	0	0	0	0	0	0	0	0	0	1	0	0	0
289	W1426_2508	0	0	0	0	0	0	0	0	0	0	0	0	0
290	W1426_2509	0	0	0	0	0	0	0	0	0	0	0	0	0
291	W1426_8579	0	0	0	0	0	0	0	0	0	0	0	0	0
292	W1426_8761	0	0	0	0	0	0	0	0	0	0	0	0	0
293	W1426_8767	0	0	0	0	0	0	0	0	0	0	0	0	0
294	W1426_8807	0	0	0	0	0	0	0	0	0	0	0	0	0
295	W1426_8817	0	0	0	0	0	0	0	0	0	0	0	0	0
296	W1426_9045	0	0	0	0	0	0	0	0	0	1	0	0	0
297	W1426_9046	0	0	0	0	0	0	0	0	0	1	0	0	0
298	W1426_5940	0	0	0	0	0	0	0	0	0	2	1	0	1
299	W1426_5941	0	0	0	0	0	0	0	0	0	2	1	0	1
300	W1426_2489	0	0	0	0	0	0	0	0	0	0	0	0	0
301	W1426_W1434	0	0	0	0	0	0	0	0	0	2	1	0	1
302	W1434_10397	0	0	0	0	2	0	0	0	0	1	0	0	0
303	W1434_10752	0	0	0	0	0	0	0	0	0	0	0	0	0
304	W1434_2503	0	0	0	0	0	0	0	0	0	0	0	0	0
305	W1434_2504	0	0	0	0	1	0	0	0	0	0	0	0	0
306	W1434_2508	0	0	0	0	0	0	0	0	0	0	0	0	0
307	W1434_2509	0	0	0	0	1	0	0	0	0	0	0	0	0
308	W1434_8579	0	0	0	0	0	0	0	0	0	0	0	0	0
309	W1434_8761	0	0	0	0	1	0	0	0	0	0	0	0	0
310	W1434_8767	0	0	0	0	1	0	0	0	0	0	0	0	0
311	W1434_8807	0	0	0	0	1	0	0	0	0	0	0	0	0
312	W1434_8817	0	0	0	0	2	0	0	0	0	0	0	0	0
313	W1434_9045	0	0	0	0	1	0	0	0	0	1	0	0	0
314	W1434_9046	0	0	0	0	1	0	0	0	0	0	0	0	0
315	W1434_5940	0	0	0	0	0	0	0	0	0	0	1	0	1
316	W1434_5941	0	0	0	0	0	0	0	0	0	0	1	0	1
317	W1434_2489	0	0	0	0	2	0	0	0	0	0	0	0	0
318	W1434_W1426	0	0	0	0	0	0	0	0	0	0	1	0	1
319	Total	6	6	35	4	27	15	7	12	9	27	7	18	15

















	A	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP
1	Table S4. Number of ir														
2	Comparison	ncra_LTR_41	ncra_LTR_71	ntet_LTR_15	ntet_LTR_39	ntet_LTR_25	ncra_LTR_49	ntet_LTR_19	ntet_LTR_05	ndisc_LTR_62	ndisc_LTR_53	ncra_LTR_76	ncra_LTR_115	ncra_LTR_38	ncra_LTR_58
285	W1426_10397	0	0	0	0	0	0	0	0	0	0	1	0	0	0
286	W1426_10752	0	0	0	0	0	1	0	0	0	0	0	0	0	0
287	W1426_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	0
288	W1426_2504	0	0	0	0	0	0	0	0	0	0	0	0	0	0
289	W1426_2508	0	0	0	0	0	0	0	0	0	0	1	0	0	0
290	W1426_2509	0	0	0	0	0	0	0	0	0	0	0	0	0	0
291	W1426_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
292	W1426_8761	0	0	0	0	0	0	0	0	0	0	0	1	0	0
293	W1426_8767	0	0	0	0	0	0	0	0	0	0	0	0	0	0
294	W1426_8807	0	0	0	0	0	1	0	0	0	0	0	0	0	0
295	W1426_8817	0	0	0	0	0	1	0	0	0	0	0	0	0	0
296	W1426_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
297	W1426_9046	0	0	0	0	0	0	0	0	0	0	0	0	0	0
298	W1426_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
299	W1426_5941	0	0	0	0	0	0	0	0	0	0	0	0	0	0
300	W1426_2489	0	0	0	0	0	0	0	0	0	0	0	1	0	0
301	W1426_W1434	0	0	0	0	0	0	0	0	0	0	0	0	0	0
302	W1434_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
303	W1434_10752	0	0	0	0	0	1	0	0	0	0	1	0	0	0
304	W1434_2503	0	0	0	0	0	0	0	0	0	0	0	1	0	0
305	W1434_2504	0	0	0	0	0	0	0	0	0	0	0	1	0	0
306	W1434_2508	0	0	0	0	0	1	0	0	0	0	0	0	0	0
307	W1434_2509	0	0	0	0	0	1	0	0	0	0	0	0	0	0
308	W1434_8579	0	0	0	0	0	1	0	0	0	0	0	0	0	0
309	W1434_8761	0	0	0	0	0	0	0	0	0	0	0	0	0	0
310	W1434_8767	0	0	0	0	0	1	0	0	0	0	1	1	0	0
311	W1434_8807	0	0	0	0	0	0	0	0	0	0	0	0	0	0
312	W1434_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
313	W1434_9045	0	0	0	0	0	4	0	0	0	0	0	0	0	0
314	W1434_9046	0	0	0	0	0	3	0	0	0	0	0	0	0	0
315	W1434_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
316	W1434_5941	0	0	0	0	0	1	0	0	0	0	0	0	0	0
317	W1434_2489	0	0	0	0	0	0	0	0	0	0	0	1	0	0
318	W1434_W1426	0	0	0	0	0	0	0	0	0	0	0	0	0	0
319	Total	4	9	6	9	1	35	3	6	2	9	7	11	4	11









	A	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD
1	Table S4. Number of ir														
2	Comparison	ndisc_LTR_54	ntet_LTR_01	ndisc_LTR_60	ncra_Gypsy_03	ncra_LTR_32	ncra_LTR_84	ntet_LTR_40	ncra_LTR_102	ncra_LTR_68	ncra_LTR_30	ncra_LTR_83	ntet_LTR_11	ntet_LTR_06	ncra_LTR_105
146	8767_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
147	8767_10752	0	0	0	0	0	0	0	0	0	0	0	0	0	0
148	8767_2503	0	1	0	0	0	0	0	0	0	0	0	0	0	0
149	8767_2504	0	1	0	0	0	0	0	0	0	0	0	0	0	0
150	8767_2508	0	0	1	0	0	0	0	0	0	0	0	0	0	0
151	8767_2509	0	0	1	0	0	0	0	0	0	0	0	0	0	0
152	8767_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
153	8767_8761	0	0	0	1	0	0	0	0	0	0	0	0	0	0
154	8767_8807	0	0	0	1	0	0	0	0	0	0	0	0	0	0
155	8767_8817	0	0	0	1	0	0	0	0	0	0	0	0	0	0
156	8767_9045	0	0	1	1	0	0	0	0	0	0	0	0	0	0
157	8767_9046	0	0	0	1	0	0	0	0	0	0	0	0	0	0
158	8767_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
159	8767_5941	0	0	0	0	0	0	0	0	0	0	0	0	0	0
160	8767_2489	0	0	1	1	0	0	0	0	0	0	0	0	0	0
161	8767_W1426	0	1	0	0	0	0	0	0	0	0	0	0	0	0
162	8767_W1434	0	1	0	0	0	0	0	0	0	0	0	0	0	0
163	8807_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
164	8807_10752	0	0	0	0	0	0	0	0	0	0	0	0	0	0
165	8807_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	0
166	8807_2504	0	0	0	0	1	0	0	0	0	0	0	0	0	0
167	8807_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	0
168	8807_2509	0	0	0	0	0	0	0	0	0	0	0	0	0	0
169	8807_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
170	8807_8761	0	0	0	0	1	0	0	0	0	0	0	0	0	0
171	8807_8767	0	0	0	0	0	1	1	0	0	0	0	0	0	0
172	8807_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
173	8807_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
174	8807_9046	0	0	0	0	0	0	0	0	0	0	0	0	0	0
175	8807_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
176	8807_5941	0	0	0	0	0	0	0	0	0	0	0	0	0	0
177	8807_2489	0	0	0	0	0	0	0	1	1	0	0	0	0	0
178	8807_W1426	0	0	0	0	0	1	0	0	0	0	0	0	0	0
179	8807_W1434	0	0	0	0	0	0	0	0	0	1	0	0	0	0







	A	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD
1	<b>Table S4. Number of ir</b>														
2	Comparison	ndisc_LTR_54	ntet_LTR_01	ndisc_LTR_60	ncra_Gypsy_03	ncra_LTR_32	ncra_LTR_84	ntet_LTR_40	ncra_LTR_102	ncra_LTR_68	ncra_LTR_30	ncra_LTR_83	ntet_LTR_11	ntet_LTR_06	ncra_LTR_105
285	W1426_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
286	W1426_10752	0	0	0	0	0	0	0	0	0	1	0	0	0	0
287	W1426_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	0
288	W1426_2504	0	0	0	0	0	0	0	0	0	0	0	0	0	0
289	W1426_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	0
290	W1426_2509	0	0	0	0	0	0	0	0	0	0	0	0	0	0
291	W1426_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
292	W1426_8761	0	0	0	0	0	0	0	0	0	0	0	0	0	0
293	W1426_8767	0	0	0	0	0	0	0	0	0	0	0	0	0	0
294	W1426_8807	0	0	0	0	0	0	0	0	0	1	0	0	0	0
295	W1426_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
296	W1426_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
297	W1426_9046	0	0	0	0	0	0	0	0	0	0	0	0	0	0
298	W1426_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
299	W1426_5941	0	0	0	0	0	0	0	0	0	0	0	0	0	0
300	W1426_2489	0	0	0	0	0	0	0	0	0	0	0	0	1	0
301	W1426_W1434	0	0	0	0	0	0	0	0	0	0	0	0	1	0
302	W1434_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
303	W1434_10752	0	0	0	0	0	0	0	0	0	0	0	0	0	0
304	W1434_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	0
305	W1434_2504	0	0	0	0	0	0	0	0	0	0	0	0	0	0
306	W1434_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	0
307	W1434_2509	0	0	0	0	0	0	0	0	0	0	0	0	0	0
308	W1434_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
309	W1434_8761	1	0	0	0	0	0	0	0	0	0	0	0	0	0
310	W1434_8767	0	0	0	0	0	0	0	0	0	0	0	0	0	0
311	W1434_8807	0	0	0	0	0	0	0	0	0	0	0	0	0	0
312	W1434_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
313	W1434_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
314	W1434_9046	0	0	0	0	0	0	0	0	0	0	0	0	0	0
315	W1434_5940	1	0	0	0	0	0	0	0	0	0	0	0	0	0
316	W1434_5941	1	0	0	0	0	0	0	0	0	0	0	0	0	0
317	W1434_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
318	W1434_W1426	0	0	0	0	0	0	0	0	0	0	0	0	0	0
319	Total	6	4	4	6	2	4	15	1	1	6	6	14	4	4















	A	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR
1	Table S4. Number of ir														
2	Comparison	ncra_LTR_110	ncra_LTR_39	ncra_LTR_85	ncra_LTR_36	ntet_LTR_07	ncra_LTR_118	ntet_Gypsy_09	ncra_LTR_06	ncra_LTR_48	ncra_LTR_103	ncra_LTR_12	ncra_LTR_56	ncra_LTR_59	ncra_LTR_37
250	5941_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
251	5941_10752	0	0	0	0	0	0	0	0	0	0	0	0	0	0
252	5941_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	0
253	5941_2504	0	0	0	0	0	0	0	0	0	0	0	0	0	0
254	5941_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	0
255	5941_2509	0	0	0	0	0	0	0	0	0	0	0	0	0	0
256	5941_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
257	5941_8761	0	0	1	0	0	0	0	0	0	0	0	0	0	0
258	5941_8767	0	0	0	0	0	0	0	0	0	0	0	0	0	0
259	5941_8807	0	0	0	0	0	0	0	0	0	0	0	0	0	0
260	5941_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
261	5941_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
262	5941_9046	0	0	0	0	0	0	0	0	0	0	0	0	0	0
263	5941_5940	0	0	0	0	1	1	1	0	0	0	0	0	0	0
264	5941_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
265	5941_W1426	0	0	1	0	0	0	0	0	0	0	0	0	0	0
266	5941_W1434	0	0	1	0	0	0	1	0	0	0	0	0	0	0
267	2489_10397	0	0	0	0	0	0	0	1	1	0	0	0	0	0
268	2489_10752	0	0	0	0	0	0	0	1	0	1	1	1	1	0
269	2489_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	1
270	2489_2504	0	0	0	1	0	0	0	0	0	0	0	0	0	1
271	2489_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	1
272	2489_2509	0	0	0	0	0	0	0	0	0	0	0	0	0	0
273	2489_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
274	2489_8579	0	0	0	0	0	0	0	1	0	0	0	1	1	0
275	2489_8761	0	0	0	0	0	0	0	0	0	0	0	1	1	1
276	2489_8767	0	0	0	0	0	0	0	1	0	1	1	0	1	0
277	2489_8807	0	0	0	0	0	0	0	0	1	0	0	0	1	1
278	2489_8817	0	0	0	0	0	0	0	1	0	0	0	0	0	0
279	2489_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
280	2489_9046	0	0	0	0	0	0	0	0	0	0	0	0	0	0
281	2489_5940	0	0	0	0	0	0	0	0	0	0	1	0	0	0
282	2489_5941	0	0	0	0	0	0	0	0	0	0	0	0	0	0
283	2489_W1426	0	0	0	0	0	0	0	0	0	0	1	0	1	0
284	2489_W1434	0	0	0	0	0	0	0	1	0	0	0	0	0	0

	A	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR
1	<b>Table S4. Number of ir</b>														
2	Comparison	ncra_LTR_110	ncra_LTR_39	ncra_LTR_85	ncra_LTR_36	ntet_LTR_07	ncra_LTR_118	ntet_Gypsy_09	ncra_LTR_06	ncra_LTR_48	ncra_LTR_103	ncra_LTR_12	ncra_LTR_56	ncra_LTR_59	ncra_LTR_37
285	W1426_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
286	W1426_10752	0	0	0	0	0	0	0	0	0	0	0	0	0	0
287	W1426_2503	0	1	0	0	0	0	0	0	0	0	0	0	0	0
288	W1426_2504	0	0	0	0	0	0	0	0	0	0	0	0	0	0
289	W1426_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	0
290	W1426_2509	0	1	0	0	0	0	0	0	0	0	0	0	0	0
291	W1426_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
292	W1426_8761	0	0	0	0	0	0	0	0	0	0	0	0	0	0
293	W1426_8767	0	1	0	0	0	0	0	0	0	0	0	0	0	0
294	W1426_8807	0	0	0	0	0	0	0	0	0	0	0	0	0	0
295	W1426_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
296	W1426_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
297	W1426_9046	0	0	0	0	0	0	0	0	0	0	0	0	0	0
298	W1426_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
299	W1426_5941	0	0	0	0	0	0	0	0	0	0	0	0	0	0
300	W1426_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
301	W1426_W1434	0	1	0	0	0	0	0	0	1	0	0	0	0	0
302	W1434_10397	0	0	0	0	0	0	0	0	0	0	0	0	0	0
303	W1434_10752	0	0	0	0	1	0	0	0	0	0	0	0	0	0
304	W1434_2503	0	0	0	0	0	0	0	0	0	0	0	0	0	0
305	W1434_2504	0	0	0	0	0	0	0	0	0	0	0	0	0	0
306	W1434_2508	0	0	0	0	0	0	0	0	0	0	0	0	0	0
307	W1434_2509	0	0	0	0	0	0	0	0	0	0	0	0	0	0
308	W1434_8579	0	0	0	0	0	0	0	0	0	0	0	0	0	0
309	W1434_8761	0	0	0	0	0	0	0	0	0	0	0	0	0	0
310	W1434_8767	0	0	0	0	0	0	0	0	0	0	0	0	0	0
311	W1434_8807	0	0	0	0	0	0	0	0	0	0	0	0	0	0
312	W1434_8817	0	0	0	0	0	0	0	0	0	0	0	0	0	0
313	W1434_9045	0	0	0	0	0	0	0	0	0	0	0	0	0	0
314	W1434_9046	0	0	0	0	0	0	0	0	0	0	0	0	0	0
315	W1434_5940	0	0	0	0	0	0	0	0	0	0	0	0	0	0
316	W1434_5941	0	0	0	0	0	0	0	0	0	0	0	0	0	0
317	W1434_2489	0	0	0	0	0	0	0	0	0	0	0	0	0	0
318	W1434_W1426	0	0	0	0	0	0	0	0	0	0	0	0	0	0
319	Total	1	8	7	3	4	2	2	6	3	2	4	3	6	5















	A	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE
1	Table S4. Number of ir													
2	Comparison	ncra_LTR_120	ncra_LTR_94	ncra_LTR_73	ncra_LTR_62	ncra_LTR_122	ncra_LTR_35	ncra_LTR_112	ncra_LTR_97	ncra_LTR_74	ncra_LTR_109	ncra_LTR_11	ncra_LTR_14	ncra_LTR_86
250	5941_10397	0	0	0	0	0	0	0	0	0	0	0	0	0
251	5941_10752	0	0	0	0	0	0	0	0	0	0	0	0	0
252	5941_2503	0	0	0	0	0	0	0	0	0	0	0	0	0
253	5941_2504	0	0	0	0	0	0	0	0	0	0	0	0	0
254	5941_2508	0	0	0	0	0	0	0	0	0	0	0	0	0
255	5941_2509	0	0	0	0	0	0	0	0	0	0	0	0	0
256	5941_8579	0	0	0	0	0	0	0	0	0	0	0	0	0
257	5941_8761	0	0	0	0	0	0	0	0	0	0	0	0	0
258	5941_8767	0	0	0	0	0	0	0	0	0	0	0	0	0
259	5941_8807	0	0	0	0	0	0	0	0	0	0	0	0	0
260	5941_8817	0	0	0	0	0	0	0	0	0	0	0	0	0
261	5941_9045	0	0	0	0	0	0	0	0	0	0	0	0	0
262	5941_9046	0	0	0	0	0	0	0	0	0	0	0	0	0
263	5941_5940	0	0	0	0	0	0	0	0	0	0	0	0	0
264	5941_2489	0	0	0	0	0	0	0	0	0	0	0	0	0
265	5941_W1426	0	0	0	0	0	0	0	0	0	0	0	0	0
266	5941_W1434	0	0	0	0	0	0	0	0	0	0	0	0	0
267	2489_10397	0	0	0	0	0	0	0	0	0	0	0	0	0
268	2489_10752	0	0	0	0	0	0	0	0	0	0	0	0	0
269	2489_2503	0	0	0	0	0	0	0	0	0	0	0	0	0
270	2489_2504	0	0	0	0	0	0	0	0	0	0	0	0	0
271	2489_2508	1	1	0	0	0	0	0	0	0	0	0	0	0
272	2489_2509	0	0	1	0	0	0	0	0	0	0	0	0	0
273	2489_5940	0	0	0	0	0	0	0	0	0	0	0	0	0
274	2489_8579	0	0	0	0	0	0	0	0	0	0	0	0	0
275	2489_8761	0	0	0	1	0	0	0	0	0	0	0	0	0
276	2489_8767	0	0	0	0	1	1	0	0	0	0	0	0	0
277	2489_8807	0	0	0	0	0	0	1	0	0	0	0	0	0
278	2489_8817	0	1	0	0	0	0	0	0	0	0	0	0	0
279	2489_9045	0	0	1	0	0	0	0	0	0	0	0	0	0
280	2489_9046	0	0	1	0	0	0	0	0	0	0	0	0	0
281	2489_5940	0	0	0	0	0	0	0	1	0	0	0	0	0
282	2489_5941	0	0	0	0	0	0	0	0	0	0	0	0	0
283	2489_W1426	0	0	0	0	0	0	0	0	0	0	0	0	0
284	2489_W1434	0	0	0	0	0	0	0	0	1	0	0	0	0

	A	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE
1	<b>Table S4. Number of ir</b>													
2	Comparison	ncra_LTR_120	ncra_LTR_94	ncra_LTR_73	ncra_LTR_62	ncra_LTR_122	ncra_LTR_35	ncra_LTR_112	ncra_LTR_97	ncra_LTR_74	ncra_LTR_109	ncra_LTR_11	ncra_LTR_14	ncra_LTR_86
285	W1426_10397	0	0	0	0	0	0	0	0	0	0	0	0	0
286	W1426_10752	0	0	0	0	0	0	0	0	0	0	0	0	0
287	W1426_2503	0	0	0	0	0	0	0	0	0	1	0	0	0
288	W1426_2504	0	0	0	0	0	0	0	0	0	0	0	0	0
289	W1426_2508	0	0	0	0	0	0	0	0	0	0	0	0	0
290	W1426_2509	0	0	0	0	0	0	0	0	0	0	0	0	0
291	W1426_8579	0	0	0	0	0	0	0	0	0	1	0	0	0
292	W1426_8761	0	0	0	0	0	0	0	0	0	0	0	0	0
293	W1426_8767	0	0	0	0	0	0	0	0	0	0	1	0	0
294	W1426_8807	0	0	0	0	0	0	0	0	0	0	1	0	0
295	W1426_8817	0	0	0	0	0	0	0	0	0	0	0	0	0
296	W1426_9045	0	0	0	0	0	0	0	0	0	1	0	0	0
297	W1426_9046	0	0	0	0	0	0	0	0	0	1	0	0	0
298	W1426_5940	0	0	0	0	0	0	0	0	0	0	1	0	0
299	W1426_5941	0	0	0	0	0	0	0	0	0	0	1	0	0
300	W1426_2489	0	0	0	0	0	0	0	0	0	0	0	0	0
301	W1426_W1434	0	0	0	0	0	0	0	0	0	0	1	0	0
302	W1434_10397	0	0	0	0	0	0	0	0	0	0	0	0	0
303	W1434_10752	0	0	0	0	0	0	0	0	0	0	0	0	0
304	W1434_2503	0	0	0	0	0	0	0	0	0	0	0	0	0
305	W1434_2504	0	0	0	0	0	0	0	0	0	0	0	0	0
306	W1434_2508	0	0	0	0	0	0	0	0	0	0	0	0	0
307	W1434_2509	0	0	0	0	0	0	0	0	0	0	0	0	0
308	W1434_8579	0	0	0	0	0	0	0	0	0	0	0	0	0
309	W1434_8761	0	0	0	0	0	0	0	0	0	0	0	0	0
310	W1434_8767	0	0	0	0	0	0	0	0	0	0	0	1	0
311	W1434_8807	0	0	0	0	0	0	0	0	0	0	0	0	0
312	W1434_8817	0	0	0	0	0	0	0	0	0	0	0	0	0
313	W1434_9045	0	0	0	0	0	0	0	0	0	0	0	0	0
314	W1434_9046	0	0	0	0	0	0	0	0	0	0	0	0	0
315	W1434_5940	0	0	0	0	0	0	0	0	0	0	1	0	0
316	W1434_5941	0	0	0	0	0	0	0	0	0	0	1	1	1
317	W1434_2489	0	0	0	0	0	0	0	0	0	0	0	0	0
318	W1434_W1426	0	0	0	0	0	0	0	0	0	0	1	0	0
319	Total	1	2	3	1	1	1	1	1	1	4	8	2	1

**Table S5.** Summary of the pairwise lineage-specific insertions found by TE subfamily in *Neurospora*. The length (bp) of the insertions per subfamily are reported together with the minimum, mean and maximum length ("Stat" column). The putative number of full-length LTR elements are indicated in the "LTR" column.

Repeat	Length	Stat	LTR
ncra_Gypsy_01	7899,7900,341,341,340,313,313,340,3	313,1011,7900	2
ncra_Gypsy_03	300,300,300,300,300,300	300,300,300	0
ncra_Gypsy_08	338,1552,1552,1552,1557,1557,1557	338,1380,1557	0
ncra_Gypsy_10	513	513,513,513	0
ncra_LTR_04	330,8533,8552,8552	330,6491,8552	3
ncra_LTR_05	7988,7988,7988,7988,7988,7926,7926	7905,7960,7988	12
ncra_LTR_06	8482,8482,8482,8482,8482,8482	8482,8482,8482	6
ncra_LTR_101	1337,1337,1337,1337,1337,1338,450,450,12	450,1020,1338	0
ncra_LTR_102	8835	8835,8835,8835	1
ncra_LTR_103	8462.8462	8462,8462,8462	2
ncra_LTR_104	724,712,762,725,724,647,574,647,712	574,705,795	0
ncra_LTR_105	6141,6141,6141,6141	6141,6141,6141	4
ncra_LTR_109	8267,8267,8267,8267	8267,8267,8267	4
ncra_LTR_11	6031,6031,1471,1471,1471,1471,1471	1471,2611,6031	2
ncra_LTR_110	8373	8373,8373,8373	1
ncra_LTR_112	8604	8604,8604,8604	1
ncra_LTR_115	6077,6061,6102,6081,6017,5989,5989	5988,6031,6102	11
ncra_LTR_116	8726,8847,8847,8848,8848,8841,8841	8726,8828,8848	7
ncra_LTR_118	342.358	342,350,358	0
ncra_LTR_12	6624,6624,6624,6624	6624,6624,6624	4
ncra_LTR_120	8293	8293,8293,8293	1
ncra_LTR_122	7921	7921,7921,7921	1
ncra_LTR_14	6090.609	6090,6090,6090	2
ncra_LTR_30	8685,8701,8701,8685,8732,8732	8685,8706,8732	6
ncra_LTR_32	7402.7402	7402,7402,7402	2
ncra_LTR_35	11351	11351,11351,11351	1
ncra_LTR_36	8307,8307,8314	8307,8309,8314	3
ncra_LTR_37	6074,6074,6074,6074,6074	6074,6074,6074	5
ncra_LTR_38	6992,6992,7106,7106	6992,7049,7106	4
ncra_LTR_39	7174,7174,7174,7174,7169,7169,7169	7165,7171,7174	8
ncra_LTR_41	7172,7172,7172,5196	5196,6678,7172	4
ncra_LTR_48	7712,7712,1148	1148,5524,7712	2
ncra_LTR_49	7781,7781,7781,7772,7774,7774,7774	487,7366,7839	33
ncra_LTR_53	854,880,876,876,872,872,876,901,876	850,1783,9473	2
ncra_LTR_56	11194,11194,11194	11194,11194,11194	3
ncra_LTR_58	7492,7492,7492,7492,7492,7492,7492	7445,7484,7504	11
ncra_LTR_59	8644,8644,8644,8644,8644,8644	8644,8644,8644	6

**Table S5.** Summary of the pairwise lineage-specific insertions found by TE subfamily in *Neurospora*. The length (bp) of the insertions per subfamily are reported together with the minimum, mean and maximum length ("Stat" column). The putative number of full-length LTR elements are indicated in the "LTR" column.

Repeat	Length	Stat	LTR
ncra_LTR_62	11296	11296,11296,11296	1
ncra_LTR_63	309,309,7477,7485,7485,7490,7490	309,5435,7490	5
ncra_LTR_64	301,301,8753,8753,301,301,320,320,8753	301,3120,8753	3
ncra_LTR_68	8729	8729,8729,8729	1
ncra_LTR_69	7555,7519,7546,7533,7555,7555,7555	7487,7533,7566	45
ncra_LTR_71	7473,7473,7473,7418,7464,7418,7473	7418,7459,7521	9
ncra_LTR_73	11442,11442,11442	11442,11442,11442	3
ncra_LTR_74	8739	8739,8739,8739	1
ncra_LTR_75	7057,7145,7145,7145,7145,7145,7106	7057,7124,7183	27
ncra_LTR_76	8251,8307,8307,8291,8291,8262,8341	8251,8292,8341	7
ncra_LTR_83	7852,7827,7827,7827,7827,7827	7827,7831,7852	6
ncra_LTR_84	8249,8249,8268,8257	8249,8255,8268	4
ncra_LTR_85	7110,7109,7109,7109,7112,7111,7111	7109,7110,7112	7
ncra_LTR_86	439	439,439,439	0
ncra_LTR_94	7034.7034	7034,7034,7034	2
ncra_LTR_97	6638	6638,6638,6638	1
ncra_Tad1_01	6930,6919,2214,6929,2030,6931,2214	384,4852,6964	0
ncra_Tad1_06	2081,1252,1798,2081,1904,1942,1904	337,1349,2331	0
ndisc_LTR_53	5932,5932,5932,5932,5932,5932,5932	5932,5932,5932	9
ndisc_LTR_54	7448,7448,7463,7441,7441,7441	7441,7447,7463	6
ndisc_LTR_60	7161,7161,7161,7161	7161,7161,7161	4
ndisc_LTR_62	449.449	449,449,449	0
ntet_Gypsy_01	393,393,393,393,393,393	393,393,393	0
ntet_Gypsy_09	999.999	999,999,999	0
ntet_LTR_01	336,336,336,336	336,336,336	0
ntet_LTR_05	7289,7431,7219,7219,7219,7219	7219,7266,7431	6
ntet_LTR_06	7036,7036,7083,7083	7036,7059,7083	4
ntet_LTR_07	7428,7428,7428,7520	7428,7451,7520	4
ntet_LTR_11	7012,6919,7012,6919,7012,6919,7012	6211,6928,7012	14
ntet_LTR_14	8189,8189,8189,8187,8187,8187	8187,8188,8189	6
ntet_LTR_15	471,471,471,471,471,471	471,471,471	0
ntet_LTR_18	7126,7126,7127,7126,7126,7126,7126	7025,7128,7203	35
ntet_LTR_19	2176,2176,2176	2176,2176,2176	0
ntet_LTR_25	469	469,469,469	0
ntet_LTR_27	2044,336,1828,1828,336,336,1828,1828	325,1318,2044	0
ntet_LTR_28	1466,1466,6085,6085	1466,3775,6085	2
ntet_LTR_35	7450,7450,7450,7450,7452,7441,7450	7441,7455,7502	9

**Table S5.** Summary of the pairwise lineage-specific insertions found by TE subfamily in *Neurospora*. The length (bp) of the insertions per subfamily are reported together with the minimum, mean and maximum length ("Stat" column). The putative number of full-length LTR elements are indicated in the "LTR" column.

Repeat	Length	Stat	LTR
ntet_LTR_37	8610,8610,8615,8608,8608,8608,8608	8608,8609,8615	15
ntet_LTR_39	8616,534,534,534,534,534,464,464,464	464,1408,8616	1
ntet_LTR_40	7154,7118,7166,7166,7166,7118,7166	7118,7147,7170	15
ntet_LTR_45	596,558,558,606,558,558,606,464,540	441,532,606	0
ntet_Tad1_01	1603,2020,1158,1158,1598,1598,1269	617,1331,2157	0
ntet_Tad1_02	6901,6901,722,616,616,616,617,617,617	610,1459,6901	0
Tad1-1	1441,617,617,932,673,673,617,617,62	352,851,1845	0
Tad3-2	1268,487,487,1380,1380,1380,1380,13	487,1492,2547	0