

Bacterial Group II Intron Genomic Neighborhoods Reflect Survival Strategies: Hiding and Hijacking

Justin Waldern,¹ Nicholas J. Schiraldi,² Marlene Belfort,^{1,3} and Olga Novikova^{*1}

¹Department of Biological Sciences and RNA Institute, University at Albany, Albany, NY

²Academic and Research Computing Center, Information Technology Services, University at Albany, Albany, NY

³Department of Biomedical Sciences, School of Public Health, University at Albany, Albany, NY

*Corresponding author: E-mail: onovikova@albany.edu.

Associate editor: Irina Arkhipova

Abstract

Group II (gII) introns are mobile retroelements that can spread to new DNA sites through retrotransposition, which can be influenced by a variety of host factors. To determine if these host factors bear any relationship to the genomic location of gII introns, we developed a bioinformatic pipeline wherein we focused on the genomic neighborhoods of bacterial gII introns within their native contexts and sought to determine global relationships between introns and their surrounding genes. We found that, although gII introns inhabit diverse regions, these neighborhoods are often functionally enriched for genes that could promote gII intron retention or proliferation. On one hand, we observe that gII introns are frequently found hiding in mobile elements or after transcription terminators. On the other hand, gII introns are enriched in locations in which they could hijack host functions for their movement, potentially timing expression of the intron with genes that produce favorable conditions for retrotransposition. Thus, we propose that gII intron distributions have been shaped by relationships with their surrounding genomic neighbors.

Key words: group II introns, retrotransposition, host functions, mobile elements.

Group II (gII) introns are mobile retroelements consisting of two active components essential for retromobility: A catalytic self-splicing RNA and a multifunctional intron-encoded protein (IEP), with reverse transcriptase, maturase, and often endonuclease activities. Upon self-splicing, the gII intron RNA forms a ribonucleoprotein particle with the IEP, which is then capable of invading DNA through two distinct retromobility pathways: retrohoming or retrotransposition (Lambowitz and Zimmerly 2011; Belfort and Lambowitz 2019). In retrohoming, gII introns invade double-stranded DNA at the cognate intron-minus site specifically and at high frequency (Cousineau et al. 1998; Smith et al. 2005). Alternatively, during retrotransposition, gII introns invade degenerate ectopic sites and target predominantly single-stranded DNA (ssDNA) at a low frequency, often at replication forks. Here, Okazaki fragments usually serve as primers for initiation of cDNA synthesis (Zhong and Lambowitz 2003; Ichiyanagi et al. 2008). Unlike retrohoming, retrotransposition allows gII introns to spread to new locations across genomic loci, contributing to the spread of these retroelements.

There are nine major phylogenetic classes of the gII introns consisting of bacterial classes A–F, mitochondrial-like (ML), and chloroplast-like classes 1 and 2 (CL1 and CL2) (Simon et al. 2009; Toro and Martinez-Abarca 2013; Toro and Nisa-Martinez 2014; Zimmerly and Semper 2015), as defined by their IEP. This classification is separate from gII intron

categorization by RNA structure into IIA, IIB, and IIC introns (Zimmerly and Semper 2015). There are some overlaps between the two classifications of introns, namely all IIC introns are also class C and vice versa. The two schemes are otherwise independent of each other (Zimmerly and Semper 2015).

Bacterial gII introns have a widespread yet patchy distribution (Dai and Zimmerly 2002; Candales et al. 2012; Toro and Martinez-Abarca 2013; Toro and Nisa-Martinez 2014). They have been described to be frequently associated with other mobile elements (Dai and Zimmerly 2002; Klein and Dunny 2002), and only occasionally interrupt important genes (Ferat et al. 2003; Chee and Takami 2005). However, little is known about the diversity of the genomic neighborhoods immediately surrounding gII introns. We became interested in the genes around gII introns following a recent study that showed how the interplay between a gII intron, resident in a conjugative plasmid, and the functionality of the gene it occupies act to mutually enhance both conjugation and retrotransposition (Novikova et al. 2014).

Here, we investigated if this interplay between the genomic neighborhood and intron biology could be a generalizable strategy for gII intron survival by analyzing the native locations of a wide variety of gII introns. We found that the functionalities of genomic neighborhoods influence colonization by gII introns, sometimes in an intron class-dependent manner. Particularly, in addition to frequently hiding in mobile

elements and after transcription terminators, gII introns are often near functions that could be hijacked, meaning that the functionality of the gene may promote retrotransposition.

Results and Discussion

Group II Intron Distribution Is Biased According to Replicon Type

Expanding upon previously established approaches for identifying gII introns (Abebe et al. 2013), we developed a pipeline for mining loci of interest from bacterial genomes by taking advantage of known semiconserved class-specific features located at 5' and 3' ends of the gII introns (fig. 1a and supplementary fig. S1, Supplementary Material online) (Candales et al. 2012; Abebe et al. 2013). In order to capture bacterial diversity while avoiding redundancy, we focused on the assembled full-length representative bacterial genomes available through the Reference Sequence (RefSeq) collection, excluding Whole Genome Shotgun sequences (O'Leary et al. 2016). With our pipeline, we identified putative 5' and 3' ends, which were in the correct 5'-3' order and appropriate distance apart (1,500–3,000 nt). We used these to assemble full-length introns, followed by extraction of corresponding IEPs and features flanking the intron (fig. 1a and supplementary fig. S1 and tables S1–S4, Supplementary Material online).

From the 1,435 RefSeq genomic sequences, we identified 863 introns from 173 bacterial species, broadly distributed among various phyla (supplementary fig. S2 and tables S1–S4, Supplementary Material online). Diversity of the gII introns in our data set was further examined using a sequence similarity network (SSN) analysis of the IEPs, which showed the presence of all known gII intron phylogenetic classes (except class A) with varying abundances (fig. 1b) (Gerlt et al. 2015; Zimmerly and Semper 2015). In particular, the SSN highlights the paucity of bacterial class A and abundance of bacterial class C introns. Indeed class C introns represent half of our mined introns—437 out of 863 or 50.6% of total introns (supplementary tables S5, Supplementary Material online).

We observed dramatically varying abundances of introns from different replicon types (e.g., chromosome or plasmid). The overwhelming majority of identified introns (93.6% or 808 introns) was located on chromosomes, with only 55 introns (6.4%) found on plasmids (supplementary tables S5 and S6, Supplementary Material online). The median number of introns we detected per chromosome was 2, in comparison with 1 intron per plasmid (fig. 1c, left and supplementary table S6, Supplementary Material online). However, some bacterial chromosomes harbor as many as 54 introns. Nevertheless, the median intron density, defined as the number of introns per each replicon sequence length, was 10 times higher in plasmids than in chromosomes, with 0.66 introns/Mb for chromosomes and 7.01 introns/Mb for plasmids (fig. 1c, right and supplementary table S6, Supplementary Material online). This finding is consistent with previous reports that gII introns retrotranspose into plasmids more frequently than into chromosomes due to the mode of

plasmid replication and the availability of ssDNA (Ichiyanagi et al. 2003).

To specifically study the relationship between the gII intron and its immediate genomic neighborhood, we considered two open reading frames flanking each side of the intron, resulting in 3,442 flanking features analyzed (fig. 1a and supplementary fig. S1 and table S3, Supplementary Material online). We chose to analyze multiple flanking genes rather than merely the host gene in order to take a broader perspective and to circumvent the often poor annotation for the features immediately adjacent to gII introns. We interrogated the functional genomic neighborhoods of gII introns based on replicon type using the Clusters of Orthologous Groups (COG) analysis, which categorizes proteins based on the biological process in which these proteins are involved (Galperin et al. 2015). We found that plasmid-based gII introns are often in neighborhoods with mobile genetic elements (MGE, COG-X) or functions involving in replication, recombination and repair (RRR, COG-L). In contrast, chromosomal gII introns show less bias in terms of neighborhood functionality, with slight biases in favor of MGE (COG-X) and RRR (COG-L) (fig. 1d and supplementary figs. S3a and S4 and tables S7–S10, Supplementary Material online). The persistence of gII introns in RRR (COG-L) and MGE (COG-X) neighborhoods on plasmids is likely a consequence of plasmids being inherently enriched for these functionalities (diCenzo and Finan 2017) in combination with the general tendency described here of gII introns to reside in COG-X and COG-L neighborhoods.

Bacterial Class C Introns Are Unique, without Biased Flanking Features

Bacterial class C introns insert after Rho-independent transcription terminators, identifying their homing site by relying more on structure than sequence specificity (Robart et al. 2007) (fig. 2b and supplementary table S11, Supplementary Material online). Interestingly, when we broke down the intron distribution of each replicon by intron class, we found that bacterial class C introns were disproportionately represented on chromosomes, constituting more than half of all chromosomal introns (fig. 1e and supplementary fig. S5 and table S5, Supplementary Material online). At the same time, when we analyzed the number of introns per class on plasmids, class C was only slightly higher than the other classes (fig. 1e and supplementary fig. S5, Supplementary Material online), consistent with class C introns not being particularly advantaged by plasmid replication. Although chromosomal localization is more favorable for class C introns than other classes, based on our COG analysis only a slight preference for MGE among flanking features was observed (fig. 2a, supplementary figs. S3b and S6 and tables S12 and S13, Supplementary Material online). This general indifference to function of neighboring genes is likely due to their unique integration sites, downstream of bacterial Rho-independent transcription terminators (Robart et al. 2007; Mohr et al. 2018).

By inserting after terminators and only rarely being transcribed, these introns limit their impact on the host by

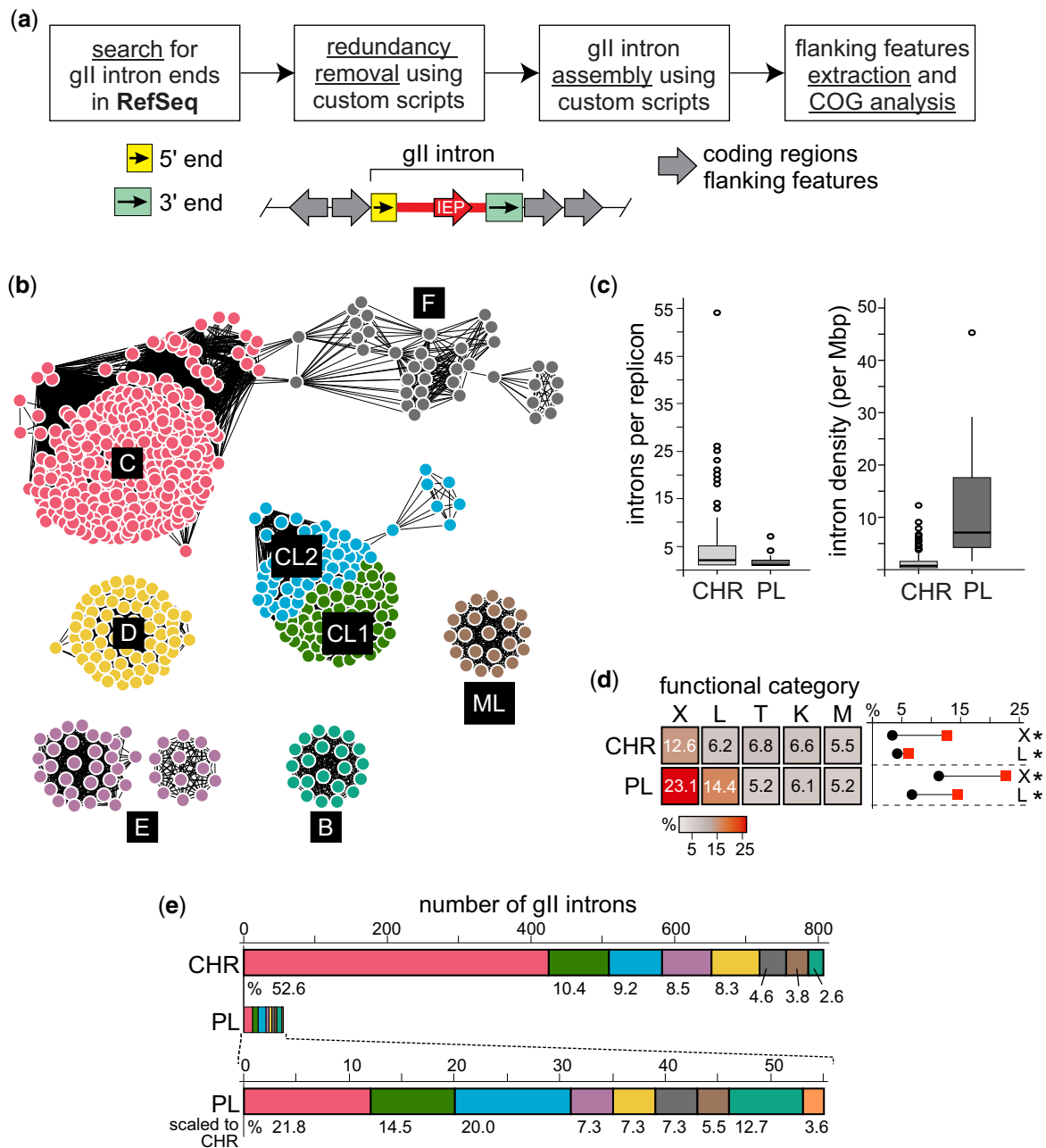


Fig. 1. Group II intron mining pipeline and results. (a) Simplified workflow for mining gII introns. Semiconserved, class-specific 5' and 3' features were used to query the RefSeq database to identify boundaries for gII introns (green and yellow boxes). These boundaries were used to locate and assemble introns (red), the corresponding intron-encoded proteins (IEPs), and two flanking features on each end (gray arrows showing orientation). Redundant (identical) introns were removed prior to retrieving flanking features. For more details, see [supplementary figure S1, Supplementary Material](#) online. (b) Sequence similarity network (SSN) of gII IEPs. IEPs identified from mining gII introns were used to produce a SSN, which clusters proteins based on degree of similarity (Gerlt et al. 2015). Each protein is represented by a node (circles) and similarity is denoted by lines connecting nodes. Clusters are colored according to the class of IEP that composes each cluster. Class A is not shown because we were unable to identify IEP sequences in the two class A introns in our data set. (c) Group II intron distributions by replicon type. Boxplots showing the number (left) and density (right, introns per megabase pair, Mb) of introns per replicon, for chromosome (CHR) or plasmid (PL). The line within the box represents the median, whereas the box represents the interquartile range. Empty circles represent outliers. Numerical values are in [supplementary table S6, Supplementary Material](#) online, with source data in [supplementary table S4, Supplementary Material](#) online. (d) Functional (COG) distribution of gII intron neighborhoods by replicon type. Heat map comparing the relative abundance of COG categories in gII intron flanking features per chromosome (CHR) and plasmid (PL). Intensity of the heatmap shows relative percentage of each category. Only top five categories are shown on the heatmap, with the full analysis in [supplementary figure S3a, Supplementary Material](#) online and a key for all COG categories in [supplementary table S7, Supplementary Material](#) online. To the right of the heatmap are plots representing the difference between relative COG abundance in gII intron neighborhoods (red squares) and the background replicon COG abundance (black circles) for COG categories X and L. Asterisks represent statistical significance at $P < 0.001$, calculated using a hypergeometric test as in [Toft et al. \(2009\)](#). COG categories shown here: X—mobile genetic elements (MGEs); L—DNA replication, recombination, and repair (RRR); T—signal transduction;

avoiding coding sequences of host genes, independently of their genomic neighborhoods (fig. 3). In addition, the structure requirement leads to a relaxed sequence specificity, allowing class C introns to proliferate more liberally to diverse locations on the rare occasions in which they are expressed, likely explaining their relative abundance on chromosomes (Robart et al. 2007). The requirement of factors beyond sequence specificity is shared with eukaryotic retrotransposons, such as Ty3, which inserts immediately upstream of PolII promoters by utilizing host proteins to identify its targets (Kirchner et al. 1995; Craig 1997). Based on their disproportionate abundance, the irrelevance of their genomic neighborhood, and their unique mechanism, class C introns have evolved a distinct survival strategy that has enabled them to become arguably the most successful class of gII introns, representing more than half of our mined introns (50.6%, fig. 1*b* and *e* and supplementary fig. S5 and table S5, Supplementary Material online).

Other gII Introns Tend to Hide in Mobile Elements

The remainder of gII intron neighborhoods demonstrated functional enrichments, both overall and in a class-dependent manner (fig. 2*a* and supplementary figs. S3 and S6 and tables S9, S12, and S13, Supplementary Material online). Overall, we frequently observed gII introns hiding in MGEs (COG-X). The most common residents of gII intron genomic neighborhoods were other gII introns, but we also often found gII introns located within a wide variety of transposases and some phage-related proteins (fig. 2*c* and supplementary tables S14 and S15, Supplementary Material online). The abundance of MGEs in gII intron neighborhoods suggests that it may be advantageous for gII introns to hide in MGEs to avoid interrupting essential host functions (fig. 3). Alternatively, it could be possible that the gII introns within a MGE can act as repressors of these elements, limiting the expression to the benefit of the host, as recently shown in Qu et al. (2018). Localizing with other MGE is not unique to gII introns, as others have proposed that retrotransposons in eukaryotes can act as “lightning rods” for new insertions of other MGEs (Jacob-Hirsch et al. 2018). Additionally, it is possible that enzymatic functions responsible for movement of the element can aid in gII intron dispersal, as is the case for host conjugative relaxase stimulating retrotransposition of the L1.LtrB gII intron (Novikova et al. 2014). This phenomenon is reminiscent of the relationship between autonomous and nonautonomous non-LTR retrotransposons (Hancks and Kazazian 2016) and composite transposons, such as Tn10, which is capable of mobilizing its internal tetracycline resistance module (Craig 1997).

By hiding on plasmids, which themselves are MGEs, gII introns can travel between bacterial cells through

conjugation. To assess the relationship between gII introns and conjugation-related neighborhoods, we searched for how often gII introns inhabit conjugative type IV secretion systems (T4SSs). When querying all fully assembled and annotated bacterial genomes, we found that 9% of conjugative T4SSs contain an associated gII intron, substantially more than the 2% found in type VI secretion systems (T6SSs), which are not involved in conjugation (supplementary table S16, Supplementary Material online). It has been shown that gII introns can capitalize on conjugative transfer for their own spread, even crossing species barriers (Belhocine et al. 2004; Belhocine et al. 2005; Novikova et al. 2014). Therefore, it appears that hiding on plasmids enables gII introns to undergo horizontal transfer and intertwines intron proliferation with that of plasmid dispersal.

Group II Introns Also Tend to Hijack Host Functions

We also observed gII introns frequently residing in neighborhoods with functionalities they could hijack (i.e., take advantage of) to promote their own proliferation, such as RRR-related proteins (COG-L) (fig. 2*a* and supplementary fig. S3*b* and tables S12, S13 and S17, S18, Supplementary Material online). RRR proteins in gII intron neighborhoods consist of a very diverse representation of helicases, ssDNA-binding proteins, and DNA polymerases (fig. 2*d*; supplementary tables S17 and S18, Supplementary Material online). RRR activity has been shown to be associated with increased gII intron retrotransposition frequency, particularly within the context of replication, where ssDNA is accessible and primers for reverse transcription are readily available (Ichihyanagi et al. 2003; Zhong and Lambowitz 2003). More recently, a gII intron was shown to interact with the β -sliding clamp of DNA polymerase III, directly linking the retrotransposition of gII introns to DNA replication (Garcia-Rodriguez et al. 2019). RRR genes are expressed at times when DNA is accessible and available for gII intron retrotransposition, suggesting that gII intron occupancy of these regions may enable the intron to capitalize on, or hijack, these functions to maximize its mobility (fig. 3) (Ichihyanagi et al. 2003). Eukaryotic retrotransposons have also been shown to integrate preferentially during DNA replication (Flasch et al. 2019), suggesting a widespread mechanism where retrotransposons proliferate by timing their expression to hijack DNA replication intermediates and functions for their mobility (fig. 3).

Furthermore, it is not unusual for an intron to be flanked by both categories X and L (fig. 2*c* and *d* and supplementary tables S14, S15 and S17, S18, Supplementary Material online), demonstrating simultaneous hiding and hijacking strategies (fig. 3, left). When we examine our data set comprehensively, we observe that the majority (57.5%) of gII introns has neighbors that are suggestive of either hiding or hijacking strategies.

FIG. 1 Continued

K—transcription; M—membrane or cell wall-related. (e) Distribution of intron classes. Shown are a comparison of the number of introns per class on chromosomes (CHR) or plasmids (PL). Colors correspond to intron classes as in (b) except for class A (not pictured in b), which is orange. Also shown is the enlarged representation of plasmid introns for comparison. Numbers below the bars represent the relative abundance (percentage) of introns detected from each class. For details, see supplementary figure S5, Supplementary Material online and supplementary table S5, Supplementary Material online.

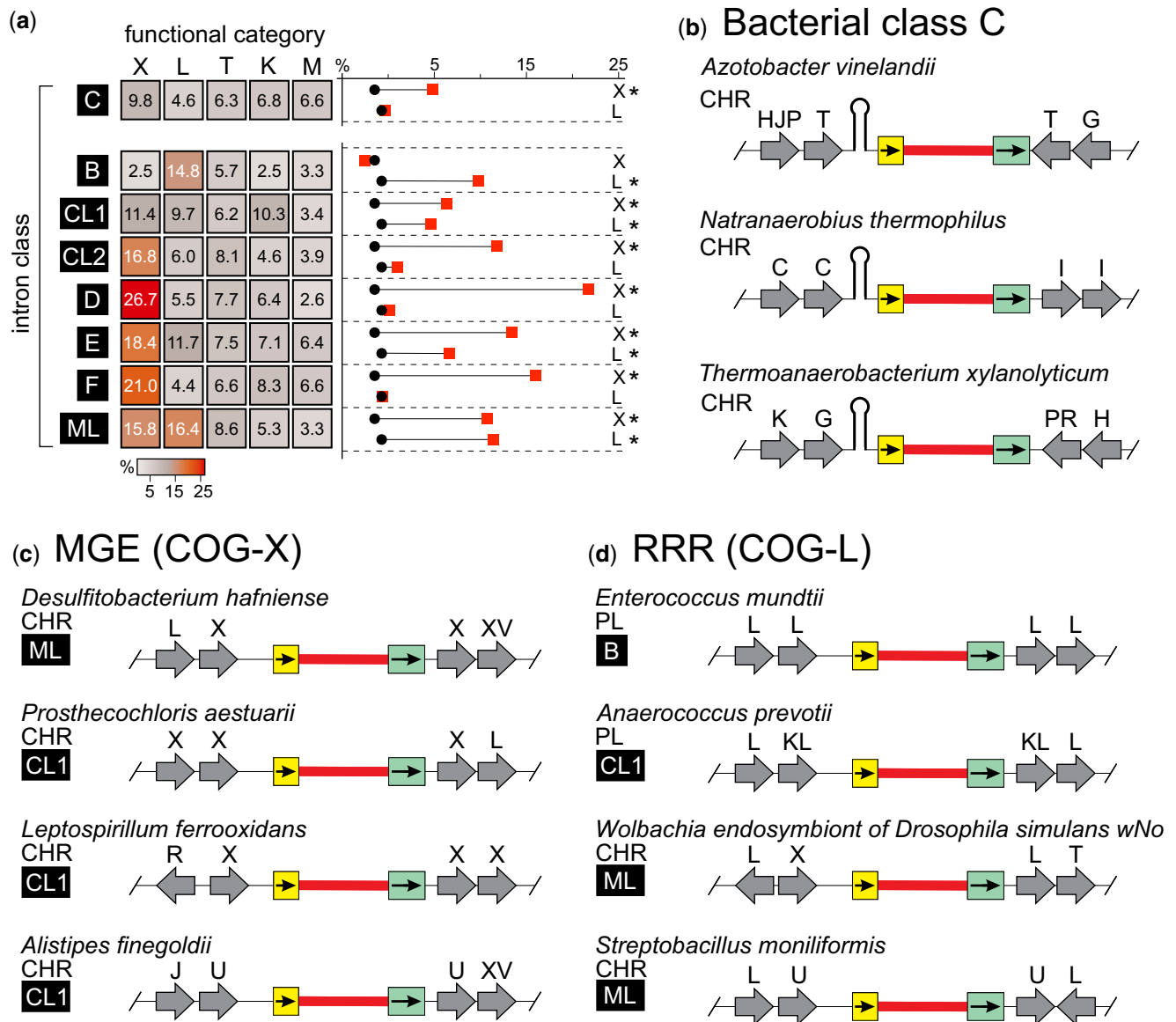


Fig. 2. Group II introns occupy distinct functional neighborhoods. (a) Group II intron neighborhood COG distribution breakdown by class, where the heatmap color represents the percent relative abundance of each category within a class. Only the top five categories are shown, but the full analysis can be found in [supplementary figure S3b, Supplementary Material](#) online and [supplementary table S12, Supplementary Material](#) online. Class A was omitted from this heatmap due to the small number of class A gII introns. To the right of the heatmap are plots representing the difference between relative COG abundance in gII intron neighborhoods (red squares) and the background COG levels from all replicons that contain gII introns (black circles). Asterisks represent statistical significance at $P < 0.001$, calculated using a hypergeometric test as in [Toft et al. \(2009\)](#). (b) Representative neighborhoods for class C introns. Class C introns show a great variety in functional neighborhoods, likely due to their mechanistic requirement for insertion after transcription terminators (stem-loop). For details on each neighborhood, see [supplementary table S11, Supplementary Material](#) online. (c) Representative neighborhoods containing mobile genetic elements (MGE) from COG-X. For details, see [supplementary table S15, Supplementary Material](#) online. (d) Representative neighborhoods from COG-L containing replication, recombination, and repair functions (RRR). For details, see [supplementary table S17, Supplementary Material](#) online. Relevant COG category key: X—mobile genetic elements (MGEs); L—DNA replication, recombination and repair (RRR); T—signal transduction; K—transcription; M—membrane or cell wall-related; C—energy production and conversion; G—carbohydrate transport and metabolism; H—coenzyme transport and metabolism; I—lipid transport and metabolism; J—translation, ribosomal structure, and biogenesis; P—inorganic ion transport and metabolism; R—general function prediction only; U—intracellular trafficking, secretion, and vesicular transport; V—defense mechanisms. All other categories are described in [supplementary table S7, Supplementary Material](#) online. If a gene is represented by more than one COG category, multiple letters are listed. All diagrams are labeled as in [figure 1a](#). The functional COG categories are labeled above gray arrows representing each feature.

In addition, these strategies are not mutually exclusive as 9% of gII introns have neighbors that represent both MGE (COG-X) and RRR (COG-L) categories ([fig. 3](#), left). An example of concomitant hiding and hijacking strategies is the

mentioned presence of a gII intron in a MGE function, conjugative relaxase, which has in turn been shown to increase retrotransposition frequencies ([Novikova et al. 2014](#)). In this example, the nicking activity of the relaxase stimulates

SURVIVAL STRATEGIES

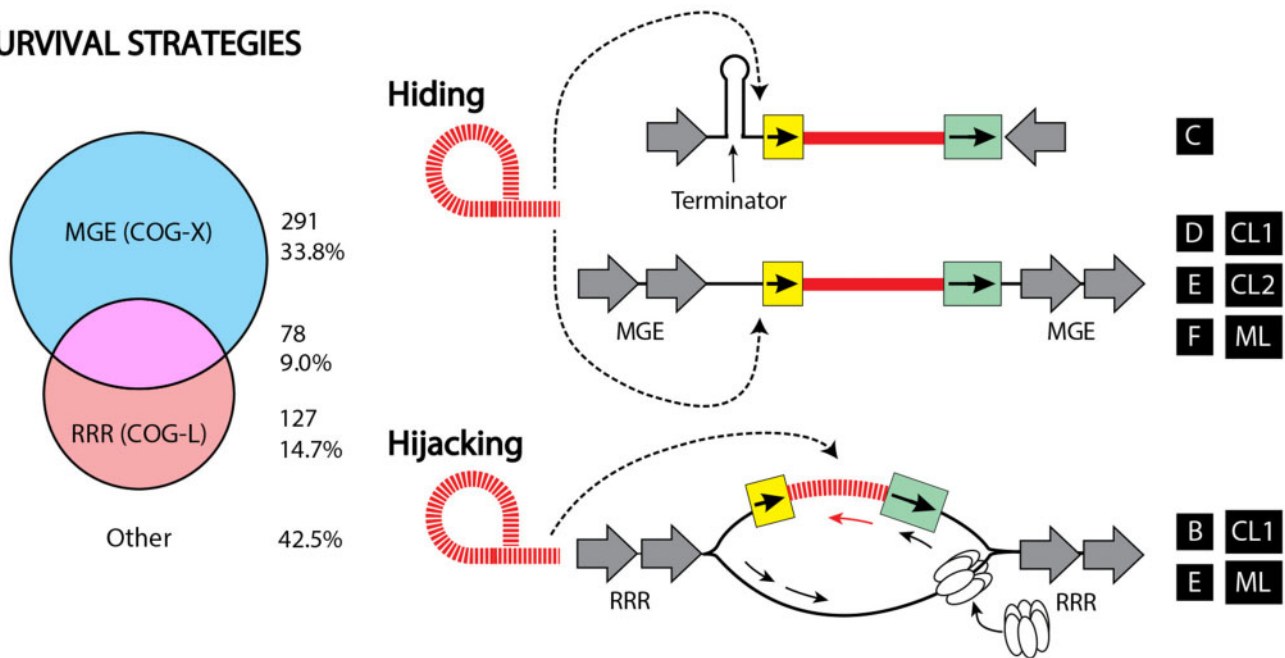


Fig. 3. Group II introns exhibit both hiding and hijacking survival strategies. Left shows a Venn diagram describing the number of gII introns containing at least one neighbor belonging to MGE (COG-X) or RRR (COG-L) categories, which are suggestive of hiding and hijacking strategies, respectively. Right shows a conceptual depiction of these strategies. Class C, the most abundant class of gII introns in bacteria, hide by inserting downstream of transcription terminators. Other classes of gII introns (CL1, CL2, D, E, F, ML) hide in genomes by avoiding key functions and colocalizing with other mobile genetic elements (MGEs). An alternative strategy that some classes utilize (B, CL1, E, ML) is to hijack the functions of their genomic neighborhoods, timing their expression with conditions that are conducive to gII intron proliferation, such as ssDNA and replication, recombination, and repair functions (RRR). Intron is in red (dashed lines RNA, solid lines DNA). Gray arrows are flanking features as in figure 2.

retrotransposition by generating accessible (nicked) DNA, which is a favorable substrate for retrotransposition. Thus, gII introns can hide in a nonessential conjugative element and simultaneously reside in a genomic neighborhood linked to a functionality, relaxase, that promotes retrotransposition. This study, conducted at a global scale, therefore anticipates more interesting examples of gII intron proliferation and persistence strategies reliant on neighborhood functionalities.

Conclusions

Genomic Neighborhoods Reflect gII Intron Survival Strategies

We have described the neighborhood environments of gII introns, as well as how these locations may contribute to their distribution, proliferation and retention (fig. 3). In particular, we note the importance not of individual genes, but of the general functionalities encoded around gII introns. Overall, we observe that gII introns tend to localize at higher densities on plasmids than chromosomes. Furthermore, they are biased to neighborhoods that may allow them to either avoid selection (hiding in MGEs or after transcription terminators) or exploit host processes for their own proliferation (hijacking RRR) (fig. 3). We propose that these observed biases in neighborhoods reflect various potential survival strategies utilized by gII introns, which may begin to aid understanding of the distribution of elements that evolved from gII introns, such as eukaryotic retrotransposons and spliceosomal introns.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) grants GM039422 and GM044844 to M.B.

References

- Abebe M, Candales MA, Duong A, Hood KS, Li T, Neufeld RAE, Shakenov A, Sun R, Wu L, Jarding AM, et al. 2013. A pipeline of programs for collecting and analyzing group II intron retroelement sequences from GenBank. *Mob DNA*. 4(1):28.
- Belfort M, Lambowitz AM. 2019. Group II intron RNPs and reverse transcriptases: from retroelements to research tools. *Cold Spring Harb Perspect Biol*. 11(4):a032375.
- Belhocine K, Plante I, Cousineau B. 2004. Conjugation mediates transfer of the Ll.LtrB group II intron between different bacterial species. *Mol Microbiol*. 51(5):1459–1469.
- Belhocine K, Yam KK, Cousineau B. 2005. Conjugative transfer of the *Lactococcus lactis* chromosomal sex factor promotes dissemination of the Ll.LtrB group II intron. *J Bacteriol*. 187(3):930–939.
- Candales MA, Duong A, Hood KS, Li T, Neufeld RA, Sun R, McNeil BA, Wu L, Jarding AM, Zimmerly S. 2012. Database for bacterial group II introns. *Nucleic Acids Res*. 40(D1):D187–D190.
- Chee GJ, Takami H. 2005. Housekeeping recA gene interrupted by group II intron in the thermophilic *Geobacillus kaustophilus*. *Gene* 363:211–220.
- Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M. 1998.

- Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* 94(4):451–462.
- Craig NL. 1997. Target site selection in transposition. *Annu Rev Biochem*. 66(1):437–474.
- Dai L, Zimmerly S. 2002. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res*. 30(5):1091–1102.
- diCenzo GC, Finan TM. 2017. The divided bacterial genome: structure, function, and evolution. *Microbiol Mol Biol Rev*. 81:e00019–17.
- Ferat JL, Le Gouar M, Michel F. 2003. A group II intron has invaded the genus *Azotobacter* and is inserted within the termination codon of the essential *groEL* gene. *Mol Microbiol*. 49(5):1407–1423.
- Flasch DA, Macia A, Sanchez L, Ljungman M, Heras SR, Garcia-Perez JL, Wilson TE, Moran JV. 2019. Genome-wide de novo L1 retrotransposition connects endonuclease activity with replication. *Cell* 177(4):837–851.e28.
- Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*. 43(D1):D261–269.
- Garcia-Rodriguez FM, Neira JL, Marcia M, Molina-Sanchez MD, Toro N. 2019. A group II intron-encoded protein interacts with the cellular replicative machinery through the beta-sliding clamp. *Nucleic Acids Res*. 47:7605–7617.
- Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim Biophys Acta*. 1854(8):1019–1037.
- Hancks DC, Kazazian HH Jr. 2016. Roles for retrotransposon insertions in human disease. *Mob DNA*. 7:9.
- Ichianagi K, Beauregard A, Belfort M. 2003. A bacterial group II intron favors retrotransposition into plasmid targets. *Proc Natl Acad Sci USA*. 100(26):15742–15747.
- Ichianagi K, Beauregard A, Lawrence S, Smith D, Cousineau B, Belfort M. 2008. Retrotransposition of the LLtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol Microbiol*. 46(5):1259–1272.
- Jacob-Hirsch J, Eyal E, Knisbacher BA, Roth J, Cesarkas K, Dor C, Farage-Barhom S, Kunik V, Simon AJ, Gal M, et al. 2018. Whole-genome sequencing reveals principles of brain retrotransposition in neurodevelopmental disorders. *Cell Res*. 28(2):187–203.
- Kirchner J, Connolly CM, Sandmeyer SB. 1995. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. *Science* 267(5203):1488–1491.
- Klein JR, Dunny GM. 2002. Bacterial group II introns and their association with mobile genetic elements. *Front Biosci*. 7(1–3):d1843–d1856.
- Lambowitz AM, Zimmerly S. 2011. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol*. 3(8):a003616.
- Mohr G, Kang SY, Park SK, Qin Y, Grohman J, Yao J, Stamos JL, Lambowitz AM. 2018. A highly proliferative group IIC intron from *Geobacillus stearothermophilus* reveals new features of group II intron mobility and splicing. *J Mol Biol*. 430(17):2760–2783.
- Novikova O, Smith D, Hahn I, Beauregard A, Belfort M. 2014. Interaction between conjugative and retrotransposable elements in horizontal gene transfer. *PLoS Genet*. 10(12):e1004853.
- O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 44(D1):D733–745.
- Qu G, Piazza CL, Smith D, Belfort M. 2018. Group II intron inhibits conjugative relaxase expression in bacteria by mRNA targeting. *Elife* 7:e34268.
- Robart AR, Seo W, Zimmerly S. 2007. Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc Natl Acad Sci USA*. 104(16):6620–6625.
- Simon DM, Kelchner SA, Zimmerly S. 2009. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. *Mol Biol Evol*. 26(12):2795–2808.
- Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M. 2005. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev*. 19(20):2477–2487.
- Toft C, Williams TA, Fares MA. 2009. Genome-wide functional divergence after the symbiosis of proteobacteria with insects unraveled through a novel computational approach. *PLoS Comput Biol*. 5(4):e1000344.
- Toro N, Martinez-Abarca F. 2013. Comprehensive phylogenetic analysis of bacterial group II intron-encoded ORFs lacking the DNA endonuclease domain reveals new varieties. *PLoS One* 8(1):e55102.
- Toro N, Nisa-Martinez R. 2014. Comprehensive phylogenetic analysis of bacterial reverse transcriptases. *PLoS One* 9(11):e114083.
- Zhong J, Lambowitz AM. 2003. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J* 22(17):4555–4565.
- Zimmerly S, Semper C. 2015. Evolution of group II introns. *Mob DNA*. 6:7.