Genomic donor cassette sharing during VLRA and VLRC assembly in jawless vertebrates

Sabyasachi Das, Emory University
Jianxu Li, Emory University
Stephen J. Holland, Max Planck Institute of Immunobiology and Epigenetics
Lakshminarayan M. Iyer, National Institutes of Health
Masayuki Hirano, Emory University
Michael Schorpp, Max Planck Institute of Immunobiology and Epigenetics
L. Aravind, National Institutes of Health
Max Dale Cooper, Emory University
Thomas Boehm, Max Planck Institute of Immunobiology and Epigenetics

Journal Title: Proceedings of the National Academy of Sciences
Volume: Volume 111, Number 41
Publisher: National Academy of Sciences | 2014-10-14, Pages 14828-14833
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1073/pnas.1415580111
Permanent URL: https://pid.emory.edu/ark:/25593/v2qs8

Final published version: http://dx.doi.org/10.1073/pnas.1415580111

Copyright information:
Freely available online through the PNAS open access option.

Accessed January 8, 2020 9:41 PM EST
Genomic donor cassette sharing during VLRA and VLRC assembly in jawless vertebrates

Sabyasachi Das1,2, Jianxu Li1,2, Stephen J. Holland1,2, Lakshminarayan M. Iyer3, Masayuki Hirano4, Michael Schorpp5, L. Aravind6,7, Max D. Cooper1,2, and Thomas Boehm1,2

*Emory Vaccine Center and Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA 30322; †Department of Developmental Immunology, Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany; and ‡National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

Contributed by Max D. Cooper, August 19, 2014 (sent for review May 27, 2014; reviewed by Jonathan P. Rast and David G. Schatz)

Lampreys possess two T-like lymphocyte lineages that express either variable lymphocyte receptor (VLR) A or VLRC antigen receptors. VLRA and VLRC+ lymphocytes share many similarities with the two principal T-cell lineages of jawed vertebrates expressing the αβ and γδ T-cell receptors (TCRs). During the assembly of VLR genes, several types of genomic cassettes are inserted, in step-wise fashion, into incomplete germ-line genes to generate the mature forms of antigen receptor genes. Unexpectedly, the structurally variable components of VLRA and VLRC receptors often possess partially identical sequences; this phenomenon of module sharing between these two VLR types occurs in both lampreys and hagfishes. By contrast, VLRA and VLRC molecules typically do not share their building blocks with the structurally analogous VLRB receptors that are expressed by B-like lymphocytes. Our studies reveal that VLRA and VLRC germ-line genes are situated in close proximity to each other in the lamprey genome and indicate the interspersed arrangement of isotype-specific and shared genomic donor cassettes; these features may facilitate the shared cassette use. The genomic structure of the VLRA/VLRC locus in lampreys is reminiscent of the interspersed nature of the TCRA/TCRD locus in jawed vertebrates that also allows the sharing of some variable gene segments during the recombinaltory assembly of TCR genes.

Significance

Lampreys possess two T-like lymphocyte lineages that express either variable lymphocyte receptor (VLR) A or VLRC antigen receptors. Despite the mutually exclusive expression pattern of VLRA and VLRC, in some cases the sequences of the two receptors are partially identical. This is the result of the shared use of genomic donor cassettes that are required to convert the incomplete VLRA and VLRC genes into functional assemblies. This feature is reminiscent of T-cell receptors of jawed vertebrates that, despite being composed of different molecular structures, also share some variable parts. The shared use of variable segments in the different antigen receptor types for T cells of all vertebrates implies a conserved functional relationship between the two prototypic T-cell lineages.


Reviewers: J.P.R., Sunnybrook Research Institute/University of Toronto; and D.G.S., Yale University School of Medicine/Howard Hughes Medical Institute.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KJ716237-KJ716286, KJ744044, KJ744045, and KJ744046).

1S.D., J.L., and S.I.H. contributed equally to this work.
2To whom correspondence may be addressed. Email: mdccoop@emory.edu or boehm@immunbio.mpg.de.

This article contains supporting information online at www.pnas.orglookup/suppl/doi:10.1073/pnas.1415580111/-/DCSupplemental.

PNAS | October 14, 2014 | vol. 111 | no. 41 | 14828-14833
whereas LRRNT and LRRCT modules were never shared. Shared use was more pronounced for LRRV modules and less frequent for LRR1 and CP modules (Fig. 1A). No module sharing was demonstrable between mature VLRA and VLRB assemblies or between mature VLRC and VLRL assemblies. Phylogenetic analyses indicated that all types of VLRA modules tend to group with VLRC modules to the exclusion of VLRB modules (SI Appendix, Figs. S1–S4). The phylogenetic trees of LRRNT (SI Appendix, Fig. S1A) and LRRCT (SI Appendix, Fig. S1B) modules showed a clear separation of the different VLR clades, with VLRA and VLRC forming distinct sister branches. By contrast, the LRR1 (SI Appendix, Fig. S2A), LRRV (SI Appendix, Figs. S3 and S4), and CP (SI Appendix, Fig. S2B) modules of mature VLRA and VLRC assemblies tended to be mixed together in the trees. This tree structure suggests an even higher degree of sequence similarity among these modules than that observed for LRRNT and LRRCT modules. Indeed, as exemplified for LRRV modules of lamprey in Fig. 1C, the majority of LRR1, LRRV, and CP modules are found in mixed groups (SI Appendix, Figs. S2–S4).

**Genomic Donor Cassette Sharing During VLRA and VLRC Assembly in Lampreys.** The germ-line *VLR* gene in lampreys is flanked by five types of donor LRR-encoding cassettes, which contribute to the assembly of mature *VLR* genes in a stepwise manner (Fig. 1A and B): 3′ LRRNT-5′ LRR1, 3′ LRR1-5′ LRRV, 3′ LRRV-5′ LRRCT, 3′ LRR1-5′ LRRV, and 3′ LRRVt-5′ LRRCT. Mismatched nucleotides are indicated in red color.

**Fig. 1.** Modules and genomic donor cassettes shared by VLRA and VLRC assemblies. (A) Schematic of the stepwise VLR assembly process, which joins incomplete genomic cassettes to complete modules. (B) Schematic of a mature VLR assembly indicating the different types of modules (Upper). VLRA and VLRC molecules of lamprey species exhibit different types of LRR modules of identical sequence; the Venn diagrams indicate the extent of module sharing for LRR1, LRRV (including LRRVt), and CP modules. (C) Phylogenetic tree of LRRV modules of lamprey VLRs. Monophyletic and strongly supported clades are collapsed into triangles, which are colored based on the type of VLR (see key); nodes supported by bootstrap values >80% are marked with red circles. Note the presence of mixed clades, containing closely related elements from different VLR isotypes. (D–G) Examples for different types of donor cassette sharing in *P. marinus*. (D) 3′ LRRNT-5′ LRR1, (E) 3′ LRR1-5′ LRRV, (F) 3′ LRRV-5′ LRRV, and (G) 3′ LRRVt-5′ LRRCT. Mismatched nucleotides are indicated in red color.

Das et al.
To assess the extent of cassette sharing, we compared the sequences of mature VLRC \((n = 70)\) and VLRA \((n = 242)\) assemblies of the sea lamprey \((Petromyzon marinus)\) with their genomic counterparts; because the nucleotide sequences of the LRRNT-, LRR1-, and LRRVt-CP-LRRCT-coding regions in VLRC are very similar in the sea lamprey and the Japanese lamprey \((Leptobranchus camtschaticum)\) \((10, 15)\), we also included a total of 100 VLRC sequences from the Japanese lamprey. In the genome assembly of \(P.\) marinus, we identified 262 genomic donor cassettes that could potentially contribute to the assembly of mature VLRA and VLRC genes, of which 53 elements are shared between VLRA and VLRC assemblies \((SI\) Appendix, Table S1). Sharing of cassettes encoding the C-terminal end of the LRRNT region is rare; of the twenty \(3\) LRRNT-5\(^{\prime}\) LRR1 cassettes, we found only one cassette that is shared between mature VLRA and VLRC \((SI\) Appendix, Fig. S6A). Of the eleven \(3\) LRR1-5\(^{\prime}\) LRRV cassettes identified in the genome assembly, 3 cassettes are shared \((SI\) Appendix, Fig. S6A). The \(3\) LRRVt-5\(^{\prime}\) LRRV donor cassettes were most frequently shared \((47/121)\) \((SI\) Appendix, Fig. S6A). Of the sixty-one \(3\) LRRVt-CP-5\(^{\prime}\) LRRCT cassettes, 2 cassettes were found to be shared \((SI\) Appendix, Fig. S6A). Mature VLRA4 and VLRC assemblies were not found to share LRRCT cassette sequences, a finding indicative of different LRRCT cassette repertoires for VLRA and VLRC \((SI\) Appendix, Fig. S1B). Although the extreme \(5\)\(^{\prime}\) ends of the genomic cassettes encoding the LRRCT modules of both VLRA and VLRC \((that\ is,\ the\ segment\ upstream\ of\ the\ sequences\ encoding\ the\ first\ cysteine\ residue\ of\ these\ modules)\) exhibit sufficient nucleotide sequence similarity to potentially accommodate the same \(3\) LRRVt-CP-5\(^{\prime}\) LRRCT cassettes \((SI\) Appendix, Fig. S5C), such hybrid sequences were not found in lamprey assemblies. Moreover, we did not find hybrid forms of VLRA assemblies containing the N-terminal part of VLRA and the C-terminal part of VLRC, or vice versa, for mature and partial VLRA and VLRC assemblies.

**Donor Cassette Sharing by Hagfish VLRA and VLRC.** Next, we examined whether the genomic donor cassette sharing also occurs between VLRA and VLRC in hagfish. Because the hagfish genome sequence is currently unavailable, a reciprocal similarity search between 60 mature VLRA, 75 VLRB, and 141 VLRC sequences of pacific hagfish \((Eptatretus stoutii)\) was performed. This analysis identified several single units of \(\geq 60\)-bp sequences having a 100% match between mature VLRA and VLRC sequences; the shared hagfish sequences that we identified are equivalent to the \(3\) LRR1-5\(^{\prime}\) LRRV, \(3\) LRR1-5\(^{\prime}\) LRRV, and \(3\) LRRVt-CP-5\(^{\prime}\) LRRCT genomic donor cassettes (each type of cassette is regarded here as a single unit) observed in lampreys \((Fig. 2A-C)\). However, we did not find identical sequences that encode the \(3\) LRRNT-5\(^{\prime}\) LRR1 region of VLRA and VLRC, except in one nonfunctional sequence \((accession\ no.\ KJ680374)\). The trend observed for donor cassette sharing for hagfish was similar to that observed for lampreys with \(3\) LRRVt-5\(^{\prime}\) LRRV donor cassettes shared most frequently; this is mirrored in the higher number of shared complete LRRV modules \((n = 26)\) compared with CP modules \((n = 4)\). A phylogenetic analysis of modules found in mature hagfish VLRC assemblies supports the notion that LRRNT and LRRCT modules form distinct clades, with VLRA and VLRC forming sister branches; notable exceptions are the LRRNTs of lamprey VLRC molecules, which did not group with LRRNTs of hagfishes \((SI\) Appendix, Fig. S4A), consistent with the unusual divergence of the lamprey VLRCs with the acquisition of a neomorph insert in LRRNT modules \((18)\). Like lamprey modules, hagfish LRR1, LRRV, and CP modules tend to be mixed together in the trees \((SI\) Appendix, Figs. S1–S4).

A comparative analysis of mature VLRA and VLRC sequences from lampreys \((192 \text{ VLRA and 60 VLRC})\) and hagfish \((60 \text{ VLRA and 142 VLRC})\) suggested that genomic donor cassette sharing for VLRA and VLRC assemblies is more frequent in hagfish than in lamprey \((SI\) Appendix, Fig. S6B). In both agnathan lineages, there was no evidence for mature, functional VLRA and VLRC modules that share LRRCT-coding cassettes; however, the finding of a nonfunctionally assembled VLRC sequence containing a VLRA-type LRRCT region encoding a CxcC motif \((accession\ no.\ BA668885)\) suggests the rare occurrence of abortive hybrid assemblies in hagfishes.

In some VLRA assemblies, the identical sequences shared by VLRA and VLRC are too long to be contributed by singlet
cassette donors; this phenomenon was observed in both hagfish (Fig. 2 D and E) and lamprey (SI Appendix, Table S1 and Fig. S7A). The extended regions of shared sequences may reflect the incorporation of multiplex donor cassettes (that is, multiples of single units present as joined elements in the germ-line) during VLRA/C assembly.

Positional Polarity of the VLR Assembly Process. The assembly of a complete variable LRR module typically involves the fusion of two 3’ LRRV-5’ LRRV donor cassettes. Combinatorial pairwise assembly of the known 121 cassettes of this type could generate up to 14,520 different LRRV modules; yet, despite the expected low probability of recurrences of the same module, many VLRA and VLRC assemblies contain identical LRRV modules. For instance, the fraction of unique sequences for the first LRRV module in VLRC assemblies ranges from 90% in Lampetra planeri to only 67% in L. camtschaticum. This phenomenon cannot be accounted for by use of germ-line-encoded fused modules alone (Fig. 2 D and E and SI Appendix, Fig. S7A and Table S1); rather it suggests a positional polarity in the assembly process with a predisposition for certain preferred pairings of specific 3’ LRRV-5’ LRRV donor cassettes. This may be due to the constraints imposed by the need for homology in the pairing of incoming with terminal cassettes. In the case of LRRV1, the sequences of genomic donor cassettes encoding this module (3’ LRR1-5’ LRRV cassettes) must be at least partly homologous to the LRR1 cassettes. Given that our analysis suggests that the LRRV modules tend to form distinct clades (SI Appendix, Fig. S4), it is likely that homology to the cassette encoding part of this module (2 of the sixty-one 3’ LRRVt-CP-5’ LRRCT cassettes) also constrains the choice of the penultimate cassette during assembly. Thus, only a subset of LRRV sequences is likely to be allowed in the first and penultimate positions, which in turn might constrain the composition of LRRVs in between them to certain preferred types. This constraint gives rise to the phenomenon of positional fidelity, whereby LRRV modules often assume the same position in the sequence of modules in different assemblies (SI Appendix, Fig. S8).

Genomic Organization of VLRA and VLRC Loci in Lampreys. The currently available genome assembly (19) provides direct evidence for the interspersion of cassettes used for either VLRA or VLRC assemblies. The VLRC locus has only two LRRCT-encoding cassettes; by contrast, numerous VLRA-specific LRRCT-encoding cassettes are found on different scaffolds; they are often flanked by shared or VLRC-specific cassettes of various kinds. Several scaffolds exhibit tandem arrays of VLRA- and VLRC-specific and shared cassettes of the same type; for instance, scaffold GL480812 features an array of 3’ LRRVt-CP-5’ LRRCT coding cassettes (Fig. 3). This

Fig. 3. Genomic organization of the sea lamprey VLRA/VLRC locus. Sequence composition of genomic scaffolds (identified by their GL designation) (19) of P. marinus, encoding VLRA- and VLRC-related sequences (not drawn to scale). Arrowheads above individual genomic components indicate their inverted orientation relative to other genomic components in a particular scaffold. The donor cassettes observed only in VLRA or VLRC assemblies are indicated by purple and green dots, respectively; donor cassettes shared by both VLRA and VLRC are indicated by orange dots. Presumptive recent duplication events are represented by different background colors for the genomic cassettes. Owing to the fragmented nature of the assembly, the order and orientation of scaffolds is arbitrary.

Das et al.

PNAS  October 14, 2014  vol. 111  no. 41  14831
type of arrangement is also present in the genome of the European brook lamprey *L. planeri* (*SI Appendix, Fig. S7 B and C*).

The hagfish *VLRC* (originally designated as *VLRA*) (12) but now reclassified as *VLRC*) and *VLRB* loci are located on the same chromosome, although far apart from each other (20). The interspersion of shared donor cassettes and those that are specifically used for *VLRA* and *VLRC* assembles suggests the possibility that the incomplete *VLRA* and *VLRC* germ-line genes may be closely linked. However, the currently available lamprey genome assemblies are not informative in this regard. In the sea lamprey genome (19), the germ-line *VLRA* gene is located in scaffold GL477382, whereas scaffold GL476420 contains the germ-line *VLRC* gene (Fig. 3 and *SI Appendix, Fig. S9A*). We found no donor LRR cassettes within the 375-kb upstream region of the *VLRC* germ-line gene of 150 kb upstream of the *VLRA* germ-line gene (*SI Appendix, Fig. S9A*). In the recently published Japanese lamprey genome (21), we found that scaffold00402 contains the germ-line *VLRA* gene, whereas the germ-line *VLRC* gene is found in scaffold00223, thus precluding definitive conclusions about the distance of *VLRA* and *VLRC* genes in the lamprey genome. To examine the possible proximity of *VLRA* and *VLRC* genes in the genome directly, genomic DNA of *L. planeri* was cleaved with rare-cutting restriction enzymes and the resulting large genomic fragments separated by pulsed-field gel electrophoresis. Southern filter hybridizations indicate that *VLRA* and *VLRC* cassettes occur on the same ~400-kb Pme fragment, but on Swa fragments of different sizes, as determined by consecutive hybridizations of the same filters with gene-specific probes (Fig. 4A and B). This finding suggests that *VLRA* and *VLRC* genes are located in close proximity in the genome of *L. planeri*. To determine whether the CP cluster (*SI Appendix, Fig. S7B*; equivalent to sequences in scaffold GL480812 of *P. marinus*) is located in the intergenic region between *VLRA* and *VLRC* genes or outside of it, the same filters were hybridized with a probe encoding the tandem array of CP sequences. The results indicate that this CP cluster occurs on a Pme fragment of different size, indicating that this sequence is not located in the intergenic region between *VLRA* and *VLRC* genes (Fig. 4C). Our analysis of the *L. planeri* genome is in general agreement with the status of the genome assemblies of the sea and Japanese lampreys and provides evidence that the donor cassettes required for the assembly of mature *VLRA* and *VLRC* genes flank the region containing the two incomplete germ-line genes.

The Evolutionary Trajectory of *VLRA* and *VLRC* Genes. The phylogenetic analysis of all VLR modules indicates that lamprey and hagfish terminal leaves generally group together with others of the same type and from the same cyclostome lineage in the trees of VLR modules (Fig. 1 and *SI Appendix, Figs. S2–S4*). For example, lamprey LRRV modules grouped with other lamprey LRRV modules, and hagfish LRRV modules grouped with hagfish LRRV modules. However, within the lamprey and hagfish lineages, sequences of different lamprey species, and sequences of the two hagfish species analyzed here grouped together (*SI Appendix, Figs. S3 and S4*). This suggested that the cassettes encoding the variable modules primarily emerged through lineage-specific expansions after the divergence of the lamprey and hagfish lineages; thereafter, they appear to have evolved to some degree in vertical fashion during speciation within the two lineages. However, prior evidence suggests that the three VLR types, including the constant parts of the genomic loci, were already present in their common ancestor (12). Thus, the diversity of sequences of donor cassettes in extant species likely emerged as the result of proliferative sweeps of particular cassettes.

Definitive evidence for such lineage-specific expansions came from sequence analysis of lamprey scaffolds containing donor VLR module genomic cassettes. These indicated duplication events involving the different types of donor cassettes (15) (Fig. 3). The presence of probable duplication events was indicated by phylogenetic analysis, a high degree of sequence similarity (≥95%) among donor cassettes that extends into the flanking sequences and matching genomic orientations. A large block duplication comprising 16 donor cassettes (twelve 3′ LRRVT′−5′ LRRV, three 3′ LRRVT′-CP-5′ LRRCT, and one 3′ LRRNT5′-LRR1) was identifiable in scaffolds GL484871 and GL480568. Tandem block duplication events involving one 3′ LRRNT-LRR1-LRRV and one 3′ LRRV-LRRV cassette were observed in close proximity of the germ-line *VLRA* gene in scaffold GL477382. Tandem block duplication events comprising four 3′ LRRVT′-CP-5′ LRRCT cassettes were observed in scaffold GL476965. Scattered events of short tandem duplication were also detected in the *VLRAC* locus.

Discussion

The *VLRA* and *VLRC* antigen receptors are expressed in mutually exclusive fashion by the two T-cell lineages of lamprey. Here, we report the surprising finding that a common repertoire of genomic cassettes is used to generate complete *VLRA* and *VLRC* assemblies not only in lampreys, but also in hagfishes. The most frequently used type of shared genomic cassette is the 3′ LRRV-5′ LRRV cassette, which encodes part of the highly variable concave antigen-binding surface of VLR molecules. This unexpected feature of cassette sharing is reminiscent of common variable (V) gene use during TCRD and TCRα gene recombination in T cells of jawed vertebrates. The *TCRα* and *TCRD* genes are interspersed, such that a particular V element situated upstream of the *TCRA/D* locus can either join to a DJα segment or to a Jα segment (22). These mutually exclusive rearrangements generate complete TCRα or TCRβ chains, which are expressed as
parts of the heterodimeric \( \alpha \) and \( \gamma \) TCR complexes of the two principal T-cell lineages. Whereas these intralocus rearrangements are part of the normal differentiation program of T cells, rare translocus rearrangements are also observed between V genes of the TCRD locus and the D\( \beta \) elements of the TCRB gene (23). By contrast, the fusion between antigen receptor genes of B cells and T cells is rare (24). The fact that in the lamprey species genomic cassettes are not shared between VLRB and VLR4 or VLRB and VLR4 genes indicates the presence of tight regulation of the assembly process. It is conceivable that this is associated with lineage-specific dichotomous regulation of the chromatin state of VLR loci, akin to the mechanisms controlling V(D)J recombination of Ig and TCR genes in jawed vertebrates (25).

Our restriction mapping experiments suggest that the VLR4 and VLRc loci are closely linked in the lamprey genome; the identification of common and unique genomic donor cassettes allowed us to construct a partial physical map of the VLRc locus in lamprey. These studies indicate that the genomic organization of the VLRc/C locus resembles that of the TCRAD locus, wherein the individual gene segments of TCR4 and TCRD genes are interspersed in a single genomic location (26–28) (SI Appendix, Fig. S9 B and C). In the TCRAD locus of jawed vertebrates, some of the variable gene segments rearrange either to form complete TCRD or TCR4 genes, whereas other variable gene segments are specific for TCR\( \alpha \) or TCR\( \delta \) chains (22). Similarly, some of the donor LRR cassettes serve a dual function in the assembly of VLR4 and VLRc genes in jawless vertebrates.

The germ-line VLR4 and VLRc genes encode only leader peptides, incomplete LRRNT, LRRCT, and the stalk regions (3, 7, 10, 12); of note, not part of the incomplete germ-line VLR4 and VLRc genes is shared or acts as donor sequence. For example, no assembled sequences were found that contain the N-terminal portion of VLR4 and the C-terminal portion of VLRc or vice versa. Given that only short sequences of similarity are required in the stepwise assembly process, such chimeric sequences should theoretically exist; however, they are either not recognized or would suggest presence of a sequence-specific mechanism distinguishing acceptor and donor sequences in VLR4 or VLRc genes, or are selected against.

The lineage-specific expansions of particular genomic donor cassettes described here suggest that, after the separation of the lamprey and hagfish lineages, these regions of the genome underwent proliferative sweeps. It is conceivable that the proliferation of certain representatives was associated with extinction of the remainder of the ancestral donor cassettes (29, 30). The extinction of ancestral modules could either have occurred by deletion of entire blocks of cassettes through homologous recombination (a process which could also result in the above-mentioned proliferation) or via gene conversion by founder sequences of the proliferating cassettes. The selective proliferation of genomic donor cassettes might reflect a lineage-specific optimization of variable modules used for recognition of distinct types of antigens. It is remarkable therefore, that despite this dynamic evolution, the sharing of a subset of modules between VLRc and VLR4 was maintained in both the hagfish and the lamprey. The genomic organization of VLRc/C locus evolved in a way that allows sharing of germ-line donor cassettes to generate the large repertoire of distinct VLR4 and VLRc assemblies in the two T-cell-like populations. Our analysis thus suggests a close functional relationship of VLR4\( \alpha \) and VLRc\( \gamma \) lymphocyte lineages in the immune system of jawless vertebrates.

Materials and Methods

The cDNA library construction and cloning were carried out as described (12, 14, 15). The identification of genomic donor cassettes and cassette sharing including the identification of potential duplication events was carried out as described (15). More details for these analyses as well as the conditions for pulsed-field gel electrophoresis and procedures for the comparison of module sequences can be found in SI Appendix, SI Materials and Methods. Animal experiments were approved by the Institutional Animal Care and Use Committee at Emory University and the Review Committee of the Max Planck Institute.

ACKNOWLEDGMENTS. This work was supported by the Max Planck Society and has received funding from the European Research Council (ERC) under the European Union’s Seventh Framework Programme (FP7/2007–2013), ERC Grant Agreement No. ERC-2012-AdG–293126. Work by L.M.L. and L.A. is funded by the Intramural Research Program of the National Institutes of Health (NIH), Department of Health and Human Services. Work by S.D., J.L., M.H., and M.D.C. was supported by NIH Grants AI072435 and GM100151 and the Georgia Research Alliance.