Antibodies

Cytidine deaminase 2 is required for VLRB antibody gene assembly in lampreys

Ryo Morimoto1, Connor P. O’Meara1, Stephen J. Holland1, Inês Trancoso1, Ahmed Souissi2, Michael Schorpp1, Danièle Vassaux2, Norimasa Iwanami1*, Orlando B. Giorgetti1, Guillaume Evanno2, Thomas Boehm†*

The antibodies of jawless vertebrates consist of leucine-rich repeat arrays encoded by somatically assembled VLRB genes. It is unknown how the incomplete germline VLRB loci are converted into functional antibody genes during B lymphocyte development in lampreys. In Lampetra planeri larvae lacking the cytidine deaminase CDA2 gene, VLRB assembly fails, whereas the T lineage–associated VLRA and VLRC antigen receptor gene assemblies occur normally. Thus, CDA2 acts in a B cell lineage–specific fashion to support the somatic diversification of VLRB antibody genes. CDA2 is closely related to activation-induced cytidine deaminase (AID), which is essential for the elaboration of immunoglobulin gene repertoires in jawed vertebrates. Our results thus identify a convergent mechanism of antigen receptor gene assembly and diversification that independently evolved in the two sister branches of vertebrates.

Introduction

Lymphocyte-based adaptive immunity arguably represents one of the major evolutionary novelties of vertebrates, which evolved more than 500 million years ago (1). Soon after, vertebrates split into several lineages, two of which survive to this day (2). About 100 extant species of lampreys and hagfishes represent the ancient lineage of jawless vertebrates (agnathans); by contrast, the group of jawed vertebrates (gnathostomes) is much more species rich, encompassing more than 50,000 species ranging from cartilaginous fishes to mammals (3). Recent work has shown that, much like the situation in the jawed vertebrate sister lineage, the immune system of jawless vertebrates is distinguished by the presence of B- and T-like cells (4, 5). However, it is unclear whether the presence of B- and T-like cell lineages is the result of convergent evolution or, alternatively, an indication that the dichotomous nature of the lymphocyte differentiation pathways was already established in the common vertebrate ancestor.

Interestingly, despite the surprising similarities at the cellular level, the molecular nature of the antigen receptors of agnathans and gnathostomes is distinctly different. The structural unit of antigen receptors of jawed vertebrates is the immunoglobulin-fold domain, which is used both for B cell receptors (BCRs) and T cell receptors (TCRs) (6). By contrast, jawless vertebrates use the leucine-rich repeat (LRR) domain as the building block for their antigen receptors, termed variable lymphocyte receptors (VLRs) (7). Notwithstanding these molecular differences, vertebrate antigen receptor genes are invariably present in incomplete form in the genome; hence, the genes must first be assembled into functional units in developing lymphocytes before they can be expressed at the cell surface (7–9).

The antibodies of jawless vertebrates consist of structurally diverse LRR-containing molecules that are encoded by somatically diversified VLRB genes (7). The antigen receptors characterizing the cellular arm of the immune system in jawless vertebrates are represented by VLRA and VLRB; these two receptors are expressed in a mutually exclusive fashion, and the two T-like lineages share certain features (4, 5). During lymphocyte development, the intervening sequences of the incomplete germline VLR genes are replaced in a stepwise manner; a gene conversion–like mechanism successively adds flanking LRR sequences present elsewhere in the genome to eventually generate a completely assembled VLR gene (7–10) (Fig. 1A and fig. S1, B and C). Because RAG-like genes—which are essential for the somatic assembly and diversification of BCRs and TCRs in jawed vertebrates (11)—were not found in lampreys, it was hypothesized that VLR genes might be assembled by an alternative mechanism, most likely by the action of cytidine deaminases (12). Studies of the expression pattern of cytidine deaminase 1 (CDA1) and CDA2, the two relevant cytidine deaminases in lampreys, suggested that the assembly of mature VLRB genes may depend on the activity of CDA2, which has sequence homology to activation-induced cytidine deaminase (AID) (4, 12). However, no definitive evidence supporting this hypothesis has been reported to date.

Results

Ontogeny of the immune system in lamprey larvae

Information about the ontogeny of the immune system of lampreys is scarce, primarily because it has proven difficult to keep in vitro fertilized lampreys in the laboratory for long periods of time. To overcome this impediment, we have established a method enabling the long-term maintenance of in vitro fertilized Lampetra planeri larvae in the laboratory (fig. S1A). This has allowed us to determine the temporal sequence of critical checkpoints in immune system development. In laboratory-reared animals, low levels of transcripts emanating from the VLRB locus (Fig. 1A) are detectable from about 6 weeks after fertilization onward (Fig. 1B), whereas the expression of assembled VLRB genes is first detectable at around 10 weeks of age (Fig. 1C). Transcripts of assembled VLRA and VLRB genes typically appear later, subsequent to transcriptional activation of their nonfunctional germline genes (Fig. 1C). Thus, in lampreys, as in jawed vertebrates (13, 14), transcriptional activity of antigen receptor gene loci precedes their...
interest because, until that time, the impact of foreign antigens on an analysis at the earliest time point of development is of particular larval larvae aged 10 to 16 weeks, here collectively referred to as young larvae. Next, we examined the characteristics of gene repertoires in conducted with animals generated and raised in 2017 and 2018. and respective time points reported here refer to experiments con- tractions from which the adult parental animals were sampled. The results contributing factor might be the genetic background of the popula- tionally defined by the detection of VLR assemblies—operated variably defined in Fig. 1A ranged from 19 to 31 amino acid residues. VLRB assemblies of 10-week-old animals are di- verse. Short (19 amino acids) and midsize (26 to 27 amino acids) sequences dominate the bimodal frequency distribution of loop lengths; only two versions of LRRCT sequences containing short loops were found, whereas the sequences of longer loops are more variable (Fig. 2A and fig. S2; details on the repertoires of individual larvae are given in tables S1 and S2). These general characteristics are also ap- parent in the VLRB repertoires of 14- and 16-week-old animals (figs. S2 and S3, A and B) and at the ~2-year stage (fig. S3C and table S1). LRRCT cassettes encoding two distinct 19–amino acid loop sequences are very frequently used at all time points (Fig. 2, A and B; fig. S2, A and B; and tables S1 and S2). We examined the possibility that the skewed representation of certain sequences in VLRB assemblies might have its origin in different copy numbers of the cor- responding LRRCT modules in the genome. To this end, we sequenced the genome of two L. planeri specimens and examined the copy numbers of several different LRRCT sequences, represent- ing the entire frequency spectrum of comple- mentary DNAs (cDNAs). Our findings suggest that the frequency of usage is independent of copy number because the examined LRRCT sequences were found to have similar copy numbers in the genomes of individuals of the same population (Fig. 2A and fig. S2, B). Among these individuals, we found, whereas the sequences of longer loops are more variable (Fig. 2A and fig. S2, B). Among these individuals, we found, whereas the sequences of longer loops are more variable. These general characteristics are also ap- parent in the VLRB repertoires of 14- and 16-week-old animals (figs. S2 and S3, A and B) and at the ~2-year stage (fig. S3C and table S1). LRRCT cassettes encoding two distinct 19–amino acid loop sequences are very frequently used at all time points (Fig. 2, A and B; fig. S2, A and B; and tables S1 and S2).

Characteristics of the VLRB gene repertoire

Next, we examined the characteristics of VLRB gene repertoires in larvae aged 10 to 16 weeks, here collectively referred to as young larvae. An analysis at the earliest time point of development is of particular interest because, until that time, the impact of foreign antigens on the structure of the repertoire is likely to have been minimal. We extracted RNA from whole animals to achieve an unbiased repres- sion of the complete repertoire. Our analysis focused on the amino acid sequences in the 5’ part of the C-terminal LRR (LRRCT) modules (Fig. 1A), which encode the highly variable loop regions of VLRB antibodies (10) that provide a substantial fraction of the binding energy in complexes of antibodies with their cognate antigens (15–18). Loop length as operationally defined in Fig. 1A ranged from 19 to 31 amino acid residues. VLRB assemblies of 10-week-old animals are diverse. Short (19 amino acids) and midsize (26 to 27 amino acids) sequences dominate the bimodal frequency distribution of loop lengths; only two versions of LRRCT sequences containing short loops were found, whereas the sequences of longer loops are more variable (Fig. 2A and fig. S2; details on the repertoires of individual larvae are given in tables S1 and S2). These general characteristics are also apparent in the VLRB repertoires of 14- and 16-week-old animals (figs. S2 and S3, A and B) and at the ~2-year stage (fig. S3C and table S1). LRRCT cassettes encoding two distinct 19–amino acid loop sequences are very frequently used at all time points (Fig. 2, A and B; fig. S2, A and B; and tables S1 and S2).

We examined the possibility that the skewed representation of certain sequences in VLRB assemblies might have its origin in different copy numbers of the corresponding LRRCT modules in the genome. To this end, we sequenced the genome of two L. planeri specimens and examined the copy numbers of several different LRRCT sequences, representing the entire frequency spectrum of complementary DNAs (cDNAs). Our findings suggest that the frequency of usage is independent of copy number because the examined LRRCT sequences were found to have similar copy numbers in the genomes of individuals of the same population (Fig. 2A and fig. S2, B). Among these individuals, we found, whereas the sequences of longer loops are more variable. These general characteristics are also apparent in the VLRB repertoires of 14- and 16-week-old animals (figs. S2 and S3, A and B) and at the ~2-year stage (fig. S3C and table S1). LRRCT cassettes encoding two distinct 19–amino acid loop sequences are very frequently used at all time points (Fig. 2, A and B; fig. S2, A and B; and tables S1 and S2).

We examined the possibility that the skewed representation of certain sequences in VLRB assemblies might have its origin in different copy numbers of the corresponding LRRCT modules in the genome. To this end, we sequenced the genome of two L. planeri specimens and examined the copy numbers of several different LRRCT sequences, representing the entire frequency spectrum of complementary DNAs (cDNAs). Our findings suggest that the frequency of usage is independent of copy number because the examined LRRCT sequences were found to have similar copy numbers in the genomes of individuals of the same population. Our analysis focused on the amino acid sequences in the 5’ part of the C-terminal LRR (LRRCT) modules (Fig. 1A), which encode the highly variable loop regions of VLRB antibodies (10) that provide a substantial fraction of the binding energy in complexes of antibodies with their cognate antigens (15–18). Loop length as operationally defined in Fig. 1A ranged from 19 to 31 amino acid residues. VLRB assemblies of 10-week-old animals are diverse. Short (19 amino acids) and midsize (26 to 27 amino acids) sequences dominate the bimodal frequency distribution of loop lengths; only two versions of LRRCT sequences containing short loops were found, whereas the sequences of longer loops are more variable (Fig. 2A and fig. S2; details on the repertoires of individual larvae are given in tables S1 and S2). These general characteristics are also apparent in the VLRB repertoires of 14- and 16-week-old animals (figs. S2 and S3, A and B) and at the ~2-year stage (fig. S3C and table S1). LRRCT cassettes encoding two distinct 19–amino acid loop sequences are very frequently used at all time points (Fig. 2, A and B; fig. S2, A and B; and tables S1 and S2).

We examined the possibility that the skewed representation of certain sequences in VLRB assemblies might have its origin in different copy numbers of the corresponding LRRCT modules in the genome. To this end, we sequenced the genome of two L. planeri specimens and examined the copy numbers of several different LRRCT sequences, representing the entire frequency spectrum of complementary DNAs (cDNAs). Our findings suggest that the frequency of usage is independent of copy number because the examined LRRCT sequences were found to have similar copy numbers in the genomes of individuals of the same population. Our analysis focused on the amino acid sequences in the 5’ part of the C-terminal LRR (LRRCT) modules (Fig. 1A), which encode the highly variable loop regions of VLRB antibodies (10) that provide a substantial fraction of the binding energy in complexes of antibodies with their cognate antigens (15–18). Loop length as operationally defined in Fig. 1A ranged from 19 to 31 amino acid residues. VLRB assemblies of 10-week-old animals are diverse. Short (19 amino acids) and midsize (26 to 27 amino acids) sequences dominate the bimodal frequency distribution of loop lengths; only two versions of LRRCT sequences containing short loops were found, whereas the sequences of longer loops are more variable (Fig. 2A and fig. S2; details on the repertoires of individual larvae are given in tables S1 and S2). These general characteristics are also apparent in the VLRB repertoires of 14- and 16-week-old animals (figs. S2 and S3, A and B) and at the ~2-year stage (fig. S3C and table S1). LRRCT cassettes encoding two distinct 19–amino acid loop sequences are very frequently used at all time points (Fig. 2, A and B; fig. S2, A and B; and tables S1 and S2).
elements (LRRV). At present, it is unclear whether the distinctly different types of VLRB assemblies are expressed by different (perhaps developmentally programmed) B cell sublineages or whether they are expressed by developmentally homogeneous but functionally distinct lymphocytes.

Genetic inactivation of \textit{CDA2} in lampreys

Genetic studies in lampreys have so far been fraught with difficulties, owing to their long generation time of several years (19) and the poor survival of in vitro fertilized animals under laboratory conditions. For this reason, it has so far proven difficult to examine gene function in the developing immune system of lampreys. Previously, it has been hypothesized, but not proven, that the formation of functional VLR genes depends on the activity of domesticated [also referred to as institutionalized (20)] cytidine deaminase (\textit{CDA}) genes of the AID-APOBEC family (12, 21, 22). We found that the expression levels of the \textit{CDA2} gene markedly increased in 10-week-old larvae, establishing a temporal coincidence of \textit{CDA2} expression, VLRB transcription, and VLRB assembly (Figs. 1, B and C, and 3A). This finding prompted us to examine the function of the \textit{CDA2} gene in lampreys by CRISPR-Cas9-mediated mutagenesis, targeting the first exon of the \textit{CDA2} gene, just upstream of the region encoding the catalytically important HxE motif (Fig. 3, B to E, and figs. S4 and S5). Seemingly unedited wild-type \textit{CDA2} sequences were found in only 2 of 28 crispants (CRISPR-Cas9-mediated mutants) but always alongside mutated \textit{CDA2} alleles (fig. S4C). We conclude that the mutagenesis is notably efficient in vivo because most of the crispants carried biallelic deleterious mutations in the \textit{CDA2} gene, in line with previous experiments targeting the sea lamprey \textit{tyrosinase} gene (23). From a practical point of view, this represents an important outcome when studying the phenotypes of recessive traits. Sequence analyses revealed the presence of variable degrees of somatic mosaicism in \textit{CDA2} crispants (Fig. 3E and fig. S4C); in some cases, it was possible to reconstruct a likely sequence of mutagenesis events during the first few cell divisions after fertilization (fig. S6). About 65% (66 of 102) of alleles recovered from \textit{CDA2} crispants exhibited frameshift mutations; most mutant alleles exhibited simple deletions, and only a minority carried insertions of nontemplated nucleotides (ranging from 1 to 14 nucleotides) (Fig. 3E and fig. S4C). All in-frame mutant alleles resulted in the net deletion of amino acid residues in the range of 3 to 23, with 4 of 33 alleles exhibiting the presence of nontemplated amino acids (fig. S5, A and B). Collectively, despite the presence of somatic mosaicism, our strategy

\textbf{Fig. 2. Characterization of the VLRB repertoire during ontogeny.} (A) Frequency of usage of loop sequences as a function of their lengths (top); number of unique loop sequences as a function of their lengths (bottom) in VLRB assemblies derived from four 10-week-old animals. (B) Frequency of usage of loop sequences as a function of their lengths (top); number of unique loop sequences as a function of their lengths (bottom) in VLRB assemblies derived from a ~2-year-old animal. (C) Frequencies of LRRCT cassettes in the genome sequences derived from sperm DNA of two animals as a function of their frequencies in expressed VLRB assemblies of young larvae. (D) Rank/frequency representation of loop sequences derived from VLRB assemblies of four 10-week-old animals; the positions of 19–amino acid–long loop sequences in the distribution are indicated in red. (E) Rank/frequency representation of loop sequences derived from VLRB assemblies of a 10-week-old animal; the positions of 19–amino acid–long loop sequences in the distribution are indicated in red. (F) Rank/rank correlation of loop sequences in VLRB assemblies of young larvae and a 2-year-old animal.

provided us with a large number of CDA2-deficient F0 animals and circumvented the need to establish stable lines before phenotypic analysis. Nonetheless, future experiments will be aimed at raising mutant fish to sexual maturity to eventually establish stable CDA2-deficient lines.

**Lack of VLRB gene assembly in CDA2-deficient larvae**

Guided by our ontogenetic analysis (Fig. 1), we examined CDA2 crispants for evidence of VLRB assembly at various time points after in vitro fertilization. In most (25 of 29) of the CDA2 crispants (age range from 16 to 26 weeks at the time of analysis), VLRB assemblies could not be detected by our sensitive reverse transcription polymerase chain reaction (RT-PCR) assays, whereas all controls (n = 6) from this series exhibited a diverse VLRB repertoire (Fig. 4A and fig. S6A). Genotyping of the few CDA2 crispants still exhibiting VLRB assemblies revealed the presence of either nonedited wild-type alleles or small in-frame deletions within a stretch of low sequence conservation as compared to the deduced protein sequences of the CDA2 genes of the sea lamprey *Petromyzon marinus* (12) and the hagfish *Eptatretus burgeri* (Fig. 3E and fig. S5C) (24). Collectively, these results establish a strong genotype/phenotype correlation between inactivation of CDA2 and the failure of VLRB assembly. In consideration of the distinctly different types of VLRB sequences with respect to loop length and sequence composition, it is worth noting that CDA2 is required for the assembly of all types of VLRB receptors in the larval repertoire. Because the transcriptional activities...
of the VLRB locus (Fig. 1B) and the CDA2 gene (Fig. 3A) coincide during lamprey development, it was important to examine the possibility that CDA2 is merely involved in activation of the VLRB locus rather than in the regulation of the VLRB assembly process itself. However, this possibility appears unlikely because the VLRB locus remains transcriptionally active even in CDA2 crisprants lacking VLRB assemblies (Fig. 4A).

**Lineage-restricted function of CDA2**

To address the question of a possible lineage-specific function of CDA2, we examined the effects of CDA2 deficiency on the assembly of the VLRA and VLRC genes that are expressed by T-lineage lymphocytes. Previously, it was established that neither VLRA nor VLRC assemblies are observed in lamprey lymphocytes expressing functionally assembled VLRB and vice versa (4, 5). We found that VLRA and VLRC assemblies occurred normally in all CDA2-deficient animals lacking VLRB assemblies (Fig. 4A). We conclude that the assembly of VLRA and VLRC genes and the assembly of the VLRB gene are genetically separable events. Collectively, our genotype/phenotype analyses establish the lineage-specific activity of CDA2 and its crucial role for the assembly of VLRB genes.

**Hematopoietic phenotype of CDA2-deficient larvae**

On the basis of the expression pattern of CDA2, VLRB-expressing lymphocytes are thought to develop in the typhlosole and the kidneys of older larvae caught in the wild (25), both of which are considered to be important sites of blood formation in lamprey at larval stages. Lack of VLRB-expressing cells in CDA2 crisprants is associated with

---

**Fig. 4. Lack of VLRB antibody gene assembly in CDA2-deficient larvae.** (A) Expression of assembled VLR genes of 26-week-old CDA2 crisprants; only animal no. 32 retains putatively functional CDA2 alleles (see Fig. 3E). M, size marker. (B) Histological sections (hematoxylin and eosin staining) depicting typhlosole (t) and kidney (k) tissues taken from CDA2 wild-type and mutant animals at 26 weeks of age. (C) Higher magnification of sections shown in (B). (D) RNA in situ hybridization with a VLRB-specific probe to locate VLRB-expressing B-like cells in the typhlosole and kidney; sections are equivalent to those shown in (B). (E) Quantification of VLRB-expressing cells in histological sections of typhlosole of control and CDA2 crispant; data (means ± SEM) for four wild-type larvae and three crispants. The difference between controls and CDA2 crisprants is significant at P < 0.0001 (Welch’s t test, two tailed).
recognizable hypocellularity in kidney and typhlosole (Fig. 4, B and C); in the typhlosole, the remaining cells have features of the erythroid and myeloid lineages, indicating that most of the hematopoietic cells in the larval typhlosole and kidney tissues are B-lineage cells. We did not detect signs of inflammation of intestinal and other tissues in CDA2 crispants; moreover, the unperturbed development of T cell lineages suggests that the thymoid—situated in the pharyngeal region (25)—functions normally. We used RNA in situ hybridization to examine the presence of VLRB-expressing cells at single-cell resolution. In wild-type larvae, VLRB-expressing cells could be easily detected; by contrast, not a single VLRB-expressing cell was detectable in these tissue sites of CDA2 crispants (Fig. 4, D and E). The absence of VLRB-expressing cells in CDA2 crispants is in line with our failure to amplify VLRB assemblies in these animals (Fig. 4A). Collectively, by using two highly sensitive assay procedures, we did not detect VLRB-expressing cells in CDA2-deficient larvae. This suggests the possibility that VLRB receptor-negative B-like cells do not survive, perhaps akin to the situation of BCR-deficient mammalian B cells (26), providing a potential explanation for the hypocellularity in the typhlosole and kidney.

DISCUSSION

Our study provides compelling evidence for a B lymphocyte lineage-specific role of CDA2 in the assembly process of lamprey VLRB antibody genes, reminiscent of the role of AID in immunoglobulin gene repertoire formation (27) and/or modification (28) in gnathostomes. Our study represents the first successful demonstration of a gene-specific interference with lamprey immune function and heralds a new era in experimental immunology of jawless vertebrates. The methodology described here is capable of providing experimentalists with sufficient numbers of VLRB-deficient animals, with which to examine the role of B-like cells in the rejection of allografts and immune responses to infectious agents. Following the paradigm established here, it should also be possible to examine whether CDA1 and its ortholog(s) are required for the assembly of functional VLRA and VLRC genes, as previously hypothesized (10) and strongly suggested by the present results. With VLRA/VLRC-deficient animals, it would be possible to examine the role of the cellular arm for immune homeostasis and antigen response.

The present results raise several additional questions. For instance, it is unclear why lamprey larvae without VLRB-expressing lymphocytes have no apparent survival disadvantage. It is conceivable that innate immune defenses, which presumably provide the required immune protection during the first 2 to 3 months in the life of lampreys, may compensate also for the loss of the humoral arm in older larvae. Nonetheless, it should be also considered that the pathogen exposure in the laboratory environment is likely lower than in the natural habitat. Our results demonstrated that CDA2 function appears dispensable for larval development and survival, supporting the notion of its exclusively tissue-specific function in the immune system.

Our results also pertain to the more general question of how self-DNA mutational processes have become incorporated into the adaptive immune systems of vertebrates. Recent work suggested a plausible molecular mechanism by which the proto-RAG protein of invertebrates has lost its ability to perform transposition and instead acquired a propensity for coupled cleavage and a preference for asymmetric DNA substrates, underpinning V(D)J recombination of BCR and TCR genes in jawed vertebrates (29, 30). Likewise, ancient cytidine deaminase mutator proteins (20, 31) may have been modified in jawless vertebrates (4, 12, 22) to now subserve cell type–specific functions in the assembly of VLR genes. The lineage-specific expression pattern of lamprey cytidine deaminases suggests one mechanism by which the potentially deleterious side effects of a self-DNA mutating enzyme could be controlled.

From a more general point of view, our results firmly establish that the evolution of domesticated versions of mutator proteins appears to be a common event contributing to the emergence of somatic diversification of antigen receptor genes as a revolutionary innovation of vertebrate immunity. The improved methodology to rear in vitro fertilized lampreys in the laboratory, combined with highly efficient CRISPR-Cas9–mediated gene disruption, opens up unprecedented opportunities to establish the genetic basis of lamprey immune functions and to further explore the general principles of vertebrate adaptive immunity.

MATERIALS AND METHODS

Animals

All animal experiments were performed in accordance with the relevant guidelines and regulations and were approved by the review committee of the Max Planck Institute of Immunobiology and Epi-genetics and the Regierungspräsidium Freiburg, Germany. Genetically modified lampreys were reared under permission of the French Ministry for Research (authorization no. 4544).

Collection of parental animals, in vitro fertilization, and maintenance of eggs and larvae

Adult brook lampreys (L. planeri) were captured by electrofishing in the Oir River (Normandy, France) and kept in the laboratory facility at 13°C until sexual maturity of females could be ascertained by skin coloration, morphology of the anal and dorsal fins, and position of ovaries and for males by a visible penis structure. In vitro fertilization (32) was carried out in a designated laboratory at the Institut National de la Recherche Agronomique (INRA) U3E unit in Rennes. Adults were anesthetized with benzocaine at a final concentration of 28 mg/liter, and gametes were collected by manual stripping. Eggs of each female were equally distributed into two petri dishes (100–mm diameter), and 20 to 30 μl of milk were added to each dish. Petri dishes were filled up to half volume with tap drinking water, which was aerated to remove chlorine traces. Thirty minutes to 1 hour after fertilization, eggs were injected with the sgRNA/Cas9 ribonucleoprotein complexes (see below). After injections, eggs (both injected and noninjected controls) were transferred to fresh water–filled petri dishes (100-mm diameter; 100 to 200 fertilized eggs per dish) and kept in a climate chamber at 12°C in the dark. After hatching, eggshells were removed. Five weeks after fertilization, larvae were transferred to 1-liter containers (four larvae per container) and kept at 12°C with a natural photoperiod, i.e., periodically adjusted to match natural conditions. Containers were supplied with a 1-cm layer of sterilized sediment (sieved to a grain size <1 mm) taken from the river of origin of the parental animals. About 90% of the water volume was changed every week, taking care not to disturb the sediment, until larvae were collected for analysis. The containers were supplied with a mixture of baker’s yeast and Spirulina microalgae twice a week. At different time points, larvae were collected, anesthetized with benzocaine at a final concentration of 28 mg/liter, and transferred either to liquid nitrogen for subsequent extraction of nucleic acids or to


6 of 9
phosphate-buffered 4% paraformaldehyde (PFA) solution for subsequent histological analysis.

**RNA/DNA extraction and cDNA synthesis**

Total RNA was isolated using TRI Reagent (Sigma), treated with ribonuclease-free deoxyribonuclease (Roche), and reextracted using TRI Reagent. About 200 ng of total RNA was reverse-transcribed to cDNA using random hexamer primers and SuperScript IV reverse transcriptase according to the manufacturer’s protocol (Thermo Fisher Scientific) in a reaction mixture of 20 μl. Genomic DNA was extracted from interphase of the TRI Reagent extractions using back extraction buffer [4 M guanidine thiocyanate, 50 mM sodium citrate, and 1 M tris (free base)].

**PCR amplification from genomic DNA**

The desired target regions of the CDA2 gene were amplified by nested PCR using gene specific primers (table S3) with the proof-reading Phusion DNA polymerase (New England Biolabs) using 1 ng of genomic DNA in a reaction mixture of 20 μl; primers are listed in table S3. Cycling conditions were as follows: initial denaturation step at 98°C for 2 min, followed by 25 cycles of denaturation (98°C for 10 s), annealing (60°C for 30 s), and elongation (72°C for 30 s). The products of the first PCR reaction were diluted to 1:50 and used as templates for a second PCR step using the same cycling conditions.

**Reverse transcription polymerase chain reaction**

cDNAs were amplified using gene specific primers (table S3) under the conditions used for amplification from genomic DNA, using 1 μl of cDNA product in a reaction mixture of 20 μl; for the second amplification step, 1 μl of a 1:50 dilution of the first reaction was used. Amplification of actin cDNA was carried out for 33 cycles.

**Quantitative PCR analyses of gene expression**

Quantitative RT-PCR was conducted using a 7500 Fast Real-Time PCR System (Applied Biosystems). Primers and probes are listed in table S3.

**Histological analyses**

At the age of 26 weeks, larvae were fixed in 4% PFA and embedded in paraffin for histological analyses. Hematoxylin and eosin staining was carried out following standard histological procedures.

**RNA in situ hybridization**

RNA in situ hybridization was performed on paraffin sections as described (25). The VLRB probe corresponds to the sequence in GenBank accession number FJ187756 (25).

**Morphometry**

Morphometric analysis was carried out on images with an Apotome instrument and AxioVs40 software (Zeiss); alternatively, ImageJ software was used.

**Whole-genome sequencing and generation of BLAST databases**

The procedures for genome sequencing and BLAST database generation used for sperm samples from Lp2 and Lp5 were described elsewhere (22). Briefly, sequencing libraries were prepared from sperm genomic DNA using the Illumina TrueSeq DNA PCR-Free LT Sample Preparation Kit (FC-121-3001); libraries were sequenced in 2 × 250–base pair paired-end mode to a depth of 150 million reads using the rapid mode on a HiSeq2500 instrument (Illumina). Paired reads were assembled into single contigs before database generation.

**sgRNA design and gene targeting**

The targeting sequence for guide RNA was designed according to instruction given by Hwang et al. (33) and cloned into the pDR274 vector. After digestion with Dra I restriction enzyme (New England Biolabs), sgRNA was generated by in vitro transcription using MAXIscript T7 Transcription Kit (Thermo Fisher Scientific). The efficiency of sgRNA target recognition was ascertained by an in vitro cleavage assay, comprising 80 ng of target PCR amplicon, 300 ng of sgRNA, and 600 ng of recombinant Cas9 protein from Streptococcus pyogenes (PNA Bio) in 20 μl of 1× CutSmart buffer (New England Biolabs). For injection into fertilized eggs, sgRNA (12.5 to 50 ng/μl), Cas9 protein (500 ng/μl), 0.05% (w/v) phenol red were mixed on ice in Danieau buffer [58 mM NaCl, 0.7 mM KCl, 0.4 mM MgSO₄, 0.6 mM Ca(NO₃)₂, and 5 mM Hepes (pH 7.6)]; about 1 to 2 nl of the solution were injected into fertilized embryos.

**Identification of CDA2 mutations**

The nature of mutations of lamprey CDA2 gene was determined by next-generation sequencing of PCR amplicons from genomic DNA templates. After column purification of PCR amplicons, the barcodes and adaptor sequences were added using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs). After individual library preparation, samples were pooled, and sequences were determined using MiSeq and subsequently deconvoluted and analyzed using standard R scripts.

**VLRB repertoire analysis by cloning**

To establish the repertoire of VLRB sequences, total RNA was extracted from the whole body of larvae. VLRB amplicons derived from 10-, 14-, and 16-week-old animals were purified using QIAGEN columns, A-tailed using Taq polymerase (Genaxxon), and then cloned into a pGEM-T Easy vector using T4 ligase (Thermo Fisher Scientific). Plasmids were purified from DH10B host bacteria using a 96-well format (NucleoSpin 96 Plasmid, MACHEREY-NAGEL). Sanger sequencing was performed using M13F primer and BigDye Terminator Cycle Sequencing Kit (Thermo Fisher Scientific).

**VLRB repertoire analysis by next-generation sequencing**

To establish the repertoire of VLRB sequences in a ~2-year-old larva, total RNA was extracted from the whole body of the entire animal. The entire RNA preparation (24 μg) was used for 40 independent cDNA synthesis reactions with 600 ng each SMARTer RACE 5′/3′ Kit (Clontech Laboratories) to label the cDNA molecules with unique molecular identifiers (UMIs) following the protocol described (34) with minor modifications; an oligo-(dT) primer was used instead of gene-specific primers, and the sequence corresponding to the partial Illumina adaptors were included in the primer sequence to avoid the subsequent ligation step (table S3). After first-strand synthesis, cDNAs were pooled and amplified with a mix of UPM_S and UPM_L primers together with the VLRB-specific primer OBG_202 (table S3). The cDNAs were reamplified with a mix of P7 + UPM_S_4N; P7 + UPM_S_5N; P7 + UPM_S_6N primers, together with a mix of VLRB-specific primers OBG_194, OBG_195, and OBG_196. After column purification of PCR amplicons, the barcodes and adaptor sequences were added using NEBNext Ultra II DNA Library Prep Kit for Illumina
(New England Biolabs); after individual library preparation, samples were pooled, and sequences were determined using MiSeq and subsequently deconvoluted and analyzed using standard R scripts. VLRB cDNA sequences were only included in the downstream analysis, if they were read at least 20 times with the same UMI and with more than one UMI barcode.

**VLRB repertoire analysis**

The number of LRR elements was determined by searching for cassettes of 24 amino acid residues conforming to the LRR consensus (7, 10). The sequences of LRRCT cassettes were extracted from the deduced protein sequences of partial VLRB assemblies by searching for the co-occurrence of NP/W/L/S and GTNTP sequence strings as boundaries; loop sequences were extracted by removing the first N-terminal nine amino acid residues of LRRCT sequences thus determined.

**Bioinformatic analyses**

The DNASTAR suite of programs, R language scripts, and Prism5 software were used for analyses of data and generation of sequence alignments and figure panels. The hagfish CDA2 sequence was retrieved from a publicly available ENSEMBL database (www.ensembl.org/index.html) using the BLASTx search algorithm.

**SUPPLEMENTARY MATERIALS**

immunology.sciencemag.org/cgi/content/full/5/45/eaba0925/DC1

Fig. S1. Schematic of VLR assembly in lampreys.

Fig. S2. Characterization of the LRRCT-encoded hypervariable loop regions of VLRB assemblies.

Fig. S3. Structure of VLRB repertoires in animals of different ages.

Fig. S4. CRISPR-Cas9-mediated mutagenesis of CDA2.

Fig. S5. Deduced protein sequences of mutant alleles in CDA2 crispants.

Fig. S6. Deduced protein sequences of partial VLRB assemblies.

Fig. S7. Structure of VLRB repertoires in animals of different ages.

Fig. S8. Characterization of the LRRCT-encoded hypervariable loop regions of VLRB assemblies.

Supplementary references can be found in the Supplementary Materials section of the paper.
Acknowledgments: We thank the Experimental Unit of Aquatic Ecology and Ecotoxicology (U3E) 1036, INRA, which is part of the research infrastructure Analysis and Experimentations on Ecosystems–France, for help with the provision and maintenance of animals. Funding: These studies received funding from the Max Planck Society, the Jung Foundation for Science and Research, and the European Research Council (ERC) under the European Union’s Seventh Framework Programme (FP7/2007-2013) through ERC grant agreement 323126 (all to T.B.). G.E. was supported by the Agence Française pour la Biodiversité. R.M. was supported by a JSPS Overseas Research Fellowship. Author contributions: R.M., C.P.O., S.J.H., I.T., A.S., M.S., D.V., N.I., O.B.G., G.E., and T.B. designed experiments. R.M., C.P.O., S.J.H., I.T., A.S., M.S., D.V., N.I., and O.B.G. performed experiments. R.M., C.P.O., S.J.H., I.T., M.S., N.I., O.B.G., and T.B. analyzed data. R.M. and T.B. wrote the manuscript. All authors discussed the results and reviewed the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: The data reported in this paper have been deposited in the NCBI Sequence Read Archive database (accession number PRJNAS565360). Additional sequences reported in this paper have been deposited in the GenBank database [accession numbers MN483451 (CDA2 locus), MN483452 to MN483574 (VLRB repertoire of 10-week-old animals), MN483575 to MN483963 (VLRB repertoire of 14-week-old animals), MN483964 to MN484595 (VLRB repertoire of 16-week-old animals), and MN484596 (VLRB germline sequence)]. All other data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials.

Submitted 4 November 2019
 Accepted 19 February 2020
 Published 13 March 2020
 doi:10.1126/sciimmunol.aba0925

Cytidine deaminase 2 is required for VLRB antibody gene assembly in lampreys
Norimasa Iwanami, Orlando B. Giorgetti, Guillaume Evanno and Thomas Boehm

DOI: 10.1126/sciimmunol.aba0925

Antibody assembly in lampreys

For B lymphocytes in jawless vertebrates to produce antibodies, a combination of VLRB gene cassettes must be stitched together to create a functional antibody gene. Circumstantial evidence based on gene expression data had previously implicated the CDA2 cytidine deaminase in this process but genetic proof was lacking. By devising techniques for CRISPR-Cas9-mediated mutagenesis of in vitro fertilized lamprey eggs and subsequent rearing of lamprey larvae in the laboratory, Morimoto et al. showed that loss-of-function mutations in CDA2 result in loss of antibody gene assembly without disrupting the formation of functional genes encoding lamprey T cell receptors. The methods pioneered in this study establish lampreys as a genetically tractable model system and will facilitate further advances in understanding immunity in this ancient group of vertebrates.

ARTICLE TOOLS
http://immunology.sciencemag.org/content/5/45/eaba0925

SUPPLEMENTARY MATERIALS
http://immunology.sciencemag.org/content/suppl/2020/03/09/5.45.eaba0925.DC1

REFERENCES
This article cites 34 articles, 7 of which you can access for free
http://immunology.sciencemag.org/content/5/45/eaba0925#BIBL

PERMISSIONS
http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service