Published in final edited form as:
Curr Opin Microbiol. 2007 August ; 10(4): 388–395. doi:10.1016/j.mib.2007.06.004.

Diversity-Generating Retroelements

Bob Medhekar[1] and **Jeff F Miller**[1,2,3,*]

1 *Department of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA*

2 *The Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA*

3 *The California Nanosystems Institute, University of California, Los Angeles, CA 90095, USA*

Summary

Parasite adaptation to dynamic host characteristics is a recurrent theme in biology. Diversity-generating retroelements (DGRs) are a newly discovered family of genetic elements that function to diversify DNA sequences and the proteins they encode. The prototype DGR was identified in a temperate bacteriophage, BPP-1, on the basis of its ability to generate variability in a gene that specifies tropism for receptor molecules on host *Bordetella* species. Tropism switching is a template-dependent, reverse transcriptase mediated process that introduces nucleotide substitutions at defined locations within a target gene. This cassette-based mechanism is theoretically capable of generating trillions of different amino acid sequences in a distal tail fiber protein, providing a vast repertoire of potential ligand-receptor interactions. Variable residues are displayed in the context of a specialized C-type lectin fold, which has evolved a unique solution for balancing protein diversity against structural stability. Homologous DGRs have been identified in the chromosomes of diverse bacterial species. These unique genetic elements have the potential to confer powerful selective advantages to their hosts, and their ability to generate novel binding specificities and dynamic antimicrobial agents suggests numerous applications.

> "Human subtlety will never devise an invention more beautiful, more simple or more direct than does Nature, because in her inventions, nothing is lacking and nothing is superfluous."
>
> - Leonardo da Vinci (1452–1519)

Diversity, the spice of life

The ability to generate adaptive diversity through stochastic mechanisms is a conserved feature of host-parasite interactions [1]. DGRs are strikingly different from previously described genetic systems that mediate phase and antigenic variation. A particularly instructive paradigm for understanding the essence of DGR function is the mammalian immune system itself. The development of antibody and T cell receptor repertoires can be conceptually divided into three stages. The first employs a germ line-encoded *diversity generator*, which generates combinatorial diversity through recombination within antibody or T cell receptor (TCR) genes [•2]. The second involves *selection* for productive events by antigen binding to B cell surface IgM or by association of a peptide-MHC complex with a TCR [3,4]. This is followed by the third stage, *amplification*, which can expand a selected lymphocyte into millions of progeny cells [3,5]. For B cells, somatic hypermutation can introduce additional variation, allowing

*Correspondence E-mail E-mail: jfmiller@ucla.edu; Tel. (+1) 310 206 7926; Fax (+1) 310 206 3865.

affinity maturation of the antibody response [6]. A remarkably similar sequence of events appears to account for the selective advantage DGRs confer to their host genomes, and the extent of diversity is impressive. The DGR in *Bordetella* phage BPP-1 can theoretically produce over 10^{14} variable nucleotide sequences at the C-terminus of the major tropism determinant protein, Mtd [••7]. In comparison, the theoretical limit of T cell receptor junctional diversity is estimated at over 10^{11} unique sequences [•2,8]. For both DGRs and lymphocyte antigen receptors, diversity generators have co-evolved with protein scaffolds that are uniquely designed to accommodate the amino acid variability required to confer distinct binding specificities.

Tropism switching by *Bordetella* bacteriophage

The infectious cycles of *Bordetella* species, which cause respiratory infections in mammals, are controlled by the BvgAS phosphorelay signal transduction system [9]. BvgAS mediates a transition between the the Bvg^+ phase, which is adapted to colonization of the respiratory tract, and the Bvg^- phase, which is adapted to *ex vivo* survival and growth in *B. bronchiseptica*, the broad host range evolutionary progenitor of *Bordetella* species that infect humans [10]. This phenotypic transition is associated with major changes in the profiles of secreted proteins and surface structures [11].

A search for generalized transducing vectors led to the discovery of a temperate phage, designated BPP-1, which displayed a marked tropism for Bvg^+ as opposed to Bvg^- phase *Bordetella* (Figure 1) [••7]. The BPP-1 receptor was identified as pertactin, a surface localized autotransporter which is only expressed in the Bvg^+ phase. Although its role during infection is controversial [12], pertactin is clearly a protective immunogen and an important component of acellular pertussis vaccines [13–15]. It was also noticed that tropism was not absolute. At a frequency of about 10^{-6}, BPP-1 variants that formed plaques with normal morphology on Bvg^- phase *Bordetella* (Bvg minus-tropic phage; BMP) were isolated, and variants that recognized both Bvg^+ and Bvg^- phase bacteria with near equal efficiency (Bvg indiscriminant *p*hage; BIP) were also obtained. Since plaque formation requires repeated rounds of replication and efficient infection of adjacent cells, these simple observations revealed that tropism switches had occurred. Electron microscopy did not detect morphologic differences between phage variants; in all cases viral particles consisted of an icosahedral head, a short tail, and six tail fibers with unusual globular structures at each distal end (see Figure 3a below) [16].

Tropism switching by *Bordetella* phage was initially viewed as an adaptation to the dynamic cell surface alterations that occur during the infectious cycles of their hosts. From the vantage point of the virus, however, the apparent objective is to generate a subpopulation of variants with the potential to recognize novel receptors. Although the BvgAS-mediated phase transition provides a convenient tool for isolating phage variants, the repertoire of receptors appears to be vast and the precise nature of the selective pressures that have given rise to tropism switching remains undefined.

The diversity generator

Comparison of the 42.5 kb dsDNA genome of BPP-1 with BMP and BIP derivatives revealed a region of variability designated the variable repeat (VR, Figure 2). VR consists of a 134 bp sequence located at the 3′ end of the *mtd* (major tropism determinant) locus [••7]. Nucleotide substitutions are always present in tropic variants, and they occur at 23 discrete positions within VR. Variability hotspots are predominantly located in the first two bases of codons, maximizing the generation of amino acid substitutions. Located downstream from *mtd* is a second copy of the 134 bp repeat, designated the template repeat (TR). In striking contrast to VR, TR is never observed to vary [••7,••17]. Adjacent to TR is the *brt* locus which encodes an enzymatically active reverse transcriptase (RT) [••7]. The presence of a RT locus in a dsDNA phage genome

was unexpected, and its proximity to VR raised the possibility that reverse transcription may play a role in generating host range variability. Located between *brt* and *mtd* is *atd* (accessory tropism determinant), a small open reading frame that encodes a 100 amino acid (aa) polypeptide. Perhaps the most remarkable observation resulting from genome inspection was that variability in VR occurred almost exclusively at positions occupied by adenine residues in their cognate TRs, and every adenine in TR (with the exception of IMH, see below) represents a potential site of variability. Addition of adenine residues in TR created new sites of VR variability, and eliminating TR adenines had the opposite effect [••17]. These observations suggested that adenine-specific mutagenesis is an inherent feature of the diversity-generating mechanism.

Additional studies revealed functional requirements for tropism switching [••7]. Electron microscopy and adhesion assays showed that *mtd* encodes the distal tail fiber protein necessary for phage infectivity, and purified Mtd binds to *Bordetella* with the same specificity as the phage from which it came [••17]. Substitution experiments in which VR sequences were "swapped" between phage with different specificities indicated that VR determines phage tropism. Null mutations in *brt* resulted in fully infective phage that had lost the ability to switch tropism, and mutations in *atd* or TR had precisely the same effect. To test the possibility that diversity generation involves a flow of information from TR to VR, synonymous substitutions were introduced into TR. As predicted, these silent nucleotide substitutions were transmitted to VR during tropism switching, along with mutations at positions corresponding to adenines. These and other results demonstrated that diversity is generated through an RT-dependent mechanism in which information is adenine-mutagenized and unidirectionally transferred from TR to VR. Mutagenesis of additional phage genes indicated that the cassette shown in Figure 2 encodes all of the loci that are required for generating diversity and it represents the prototypical DGR.

Diversity-generation is characterized by mutagenesis and directional transfer to a specific targeted region of DNA. What might determine the directionality of "mutagenic homing?" At the 3′ ends of TR and VR are identical 14 bp segments consisting of G and C residues, followed by 21 bp sequences which differ at 5 positions between the two repeats. Substitution of the 21 bp VR sequence with the corresponding one from TR eliminated mutagenic homing. The reverse substitution, in which the 3′ end of TR was engineered to be identical to its counterpart from VR, maintained tropism switching and gained the ability to diversify TR, an event that is never observed in wild type phage [••17]. These observations demonstrate that the sequence designated IMH (initiation of mutagenic homing) determines the directional transfer of sequence information. IMH* designates the corresponding sequence located downstream from TR (Figure 2). Although wild type phage can presumably undergo unlimited rounds of tropism switching, phage with identical IMH sequences at the 3′ ends of TR and VR should eventually *lose* the ability to diversify. Any time an adenine in TR is substituted with another nucleotide, it will lose its ability to diversify the corresponding site in VR. Thus, there appears to be a simple yet elegant "genetic logic" that underlies DGR function. TR, which contains the master copy of the information required for mutagenic homing is never corrupted, and the capacity to diversify is forever preserved.

Although progress has been made in defining genetic requirements for DGR function, the biochemical mechanism responsible for diversity generation has yet to be determined. In recent experiments in which a group I intron was inserted into TR, mutagenic homing resulted in the acquisition of precisely spliced exon sequences by the cognate VR (*Guo et al.*, unpublished). These observations establish the involvement of an RNA intermediate, but where it begins, where it ends, and what might control its expression are unknown. Similarly, the precise role of IMH in mutagenic homing has not been defined, the function of Atd has yet to be determined, and the mechanism of adenine mutagenesis remains elusive. The models in Figure 2 illustrate

an array of working hypotheses that are consistent with available data and make testable predictions for future studies.

"Ordered disorder" – the C-type lectin scaffold

If sequence variation is to confer a selective advantage, a diversity generator must interface with a protein scaffold that can display amino acid variability while maintaining structural stability. Until the discovery of DGRs, the immunoglobulin (Ig)-fold found in antibodies and T cell receptors provided the major paradigm for how this might occur [18]. Crystal structures of 5 Mtd variants representing different receptor specificities were determined at 1.56–2.52 Å resolution [••19]. Tropic variants were nearly identical in overall structure (Figure 3a–d), forming tetrahedral-shaped trimers which localize to the distal ends of tail fibers (*Zhou et al.*, unpublished) [16]. A large hydrophobic interface stabilizes the trimer, and each 381 aa subunit is organized into three discrete domains (Figure 3b). At the N-terminus (aa 1–48) is a β-prism, which forms the threefold symmetric apex of the pyramid. This is connected by a short, intertwined 3_{10}α-helix to a β-sandwich (aa 54–170) which consists of a series of three- and four-stranded anti-parallel β-sheets. The third domain, at the C-terminus of Mtd (aa 171–381), presents VR-encoded variable residues on a ligand binding surface using a fold that is highly characteristic of a superfamily of proteins classified as C-type lectins [••19]. C-type lectins are extremely widespread in nature, they bind sugar as well as protein ligands, and they perform a diverse array of functions ranging from cell-cell adhesion to pathogen recognition by components of the innate immune system [20–•22]. The CLec-fold in Mtd is characterized by a two stranded anti parallel β-sheet formed by the N- and C-termini of the domain, connected by two α-helices to a three-stranded antiparallel β-sheet (Figure 3b). Mtd contains two additional inserts that help stabilize the variable region, and the extended β2-β 3 loop from one subunit inserts into a pocket on a neighboring subunit to help stabilize VR.

Diversified adenines in TR correspond to 12 variable amino acid residues in the VR-encoded C-terminus of Mtd. These are organized into a discrete, solvent exposed receptor-binding site on the external face of the β3 β4 β4′ sheet. Each Mtd trimer presents three VR surfaces located on the bottom face of the complex (Figure 3a). Co-localization and surface exposure of non-contiguous variable residues illustrates co-evolution between the placement of variable adenines in TR and the physical structure of the diversified protein. All but two of the variable residues in VR are encoded by AAC "codons" in TR [••7]. Adenine-specific mutagenesis of AAC permits substitution by 14 amino acids covering the entire range of chemical character, and the use of AAC rules out the possibility of a nonsense codon being introduced. Figure 3d shows the main chain conformation and variable side chains of five VR sequences superimposed on each other [••19]. These sequences recognize different receptors and they differ dramatically in chemical character, yet the backbone structure is essentially invariant. This represents a remarkable feature of Mtd and a major contrast with immunoglobulins, which display enormous conformational flexibility in their antigen binding sites. Presumably, the Clec fold in Mtd has evolved to maximize protein stability while displaying amino acid diversity.

Diversity generating retroelements are widespread in bacterial genomes

The ability to diversify protein domains involved in ligand-receptor interactions has the potential to confer a powerful selective advantage. Using the *Bordetella* phage DGR as a signature, homologous retroelements have been identified in nearly 30 genomes representing diverse bacterial species (*Zimmerly S et al.*, personal communication, *Gingery M et al.*, unpublished data) [••17]. The examples in Figure 4a include DGRs predicted in the genomes of a *V. harveyi* phage, a *M. magneticum* phage, and several putative elements that do not appear to be phage associated or recently acquired through horizontal gene transfer. These include

DGRs in the genomes of photosynthetic marine cyanobacteria (*T. erythraeum*), green sulfur bacteria (*C. chlorochromatii*), and anaerobic spirochetes implicated in the development of periodontal disease (*T. denticola*).

All of the elements in Figure 4a are predicted to encode related RTs. They also include repeated sequences analogous to TR and VR in the BPP-1 DGR. Although TR and VR analogs typically display little or no sequence similarity, TR/VR pairs differ almost exclusively at sites corresponding to adenines in TR. This suggests that adenine-specific mutagenesis is an inherent feature of DGR function. DGR-associated RTases comprise a distinct family with unique polymorphisms that may reflect unusual enzymatic specificities (*Hodes A et al.*, unpublished data, Figure 4b) [••17]. Based on the number of adenines in their TRs, the theoretical diversity that could be generated by the elements in Figure 4a ranges from 10^{14} (BPP-1) to 10^{24} nucleotide sequences (*T. denticola*). Although ORFs that are predicted to be diversified display minimal sequence identity, VR analogs are always located at their extreme C-terminal ends. Profile-based alignments predict that the C-terminal domains encoded by VR-containing ORFs adopt Clec folds similar to the one in Mtd (*Gingery M et al.*, unpublished data) [••19]. Comparative sequence analysis supports the hypothesis that both the genetic mechanism that generates diversity, and the protein scaffold used to display it, are conserved features of all DGRs.

Although DGRs are likely to function in a similar manner, they display a considerable amount of diversity in their organization. In the examples shown in Figure 4a, TR sequences are located upstream, downstream, or within RTase coding sequences. There is also an apparent dichotomy between the presence of *atd* homologs and loci encoding HRDC (Helicase and RNAseD C-terminal) domains. Most interesting, however, is the observation that DGRs may be capable of diversifying multiple linked (*M. magneticum*) or unlinked (*T. denticola*) VR-containing loci. In *T. denticola*, the N-termini of diversified proteins contain lipoprotein consensus sequences which predict lipid modification of the mature protein, signal peptide cleavage by Lsp, and localization to the outer membrane [23,24]. An intriguing possibility is that the *T. denticola* DGR system may be responsible for the surface display of a family of diversified proteins.

Conclusions

It is remarkable to consider that over 40% of the human genome is comprised of retroelements and degraded derivatives of them [25]. Retrotransposons, retroviruses, non-LTR retroelements, group II introns and related elements share two fundamental characteristics. First, they replicate through obligatory RNA intermediates using reverse transcriptase-dependent mechanisms [25–•27,28]. Second, they do not appear to confer an *obvious* selective advantage to their hosts; they are neutral at best and can be overtly deleterious. Although the issue provokes debate, in an evolutionary sense retroelements are often thought of as genetic parasites, or “selfish DNA”, and mobility may be a necessary consequence of their parasitic nature [28–30]. In contrast, the *Bordetella* phage DGR confers a powerful selective advantage to its host and the same is likely to be true for all DGRs. It has been proposed that the adaptation of mobility to generate protein variability has allowed DGRs to evolve as stable and beneficial components of host genomes [••17].

Diversity-generating retroelements were first reported in 2002 and many questions regarding their function remain to be addressed. What might account for the adenine-specific mutagenesis that characterizes these elements and what is the biochemical basis of directional homing? Are differences in DGR structure reflective of differences in their function? How extensive is the repertoire of ligands that can be recognized by diversified proteins and what are the “rules” of recognition? Although the *Bordetella* phage DGR has been studied in some detail, the selective advantages conferred by chromosomal DGRs present in diverse bacterial species have yet to

be determined. In addition to their likely importance for the organisms that encode them, DGRs may have practical applications as self-contained diversity generators. How “plastic” are they and can they be adapted to diversify heterologous proteins to evolve binding specificities, enzymatic functions, or other properties of interest? The presence of DGRs in phage genomes is of particular interest. Do phage that encode DGRs function as dynamic antimicrobial agents adapted to overcome resistance by their hosts? The possibility that DGR-endowed phage could be harnessed to eliminate pathogens in the environment, or to treat antibiotic-resistant infections, warrants further attention.

Acknowledgments

We would like to thank David Martin, M.D. and Partho Ghosh, Ph.D. for their critical reading of the manuscript. B.A.M was supported by the UCLA Medical Scientist Training Program training grant GM 08042. This work was supported by NIH Grants AI 61598 and AI 71204, and UC Discovery Grant Bio04-10428 to J.F.M.

References

1. van der Woude MW, Baumler AJ. Phase and antigenic variation in bacteria. Clin Microbiol Rev 2004;17:581–611. [PubMed: 15258095]

2•. Nemazee D. Receptor editing in lymphocyte development and central tolerance. Nat Rev Immunol 2006;6:728–740. [PubMed: 16998507]This review comprehensively examines the generation of diverse lymphocyte antigen receptor specificities and pays particular attention to the role of receptor editing as a central tolerance mechanism

3. Thomas MD, Srivastava B, Allman D. Regulation of peripheral B cell maturation. Cell Immunol 2006;239:92–102. [PubMed: 16797504]

4. Aliahmad P, Kaye J. Commitment issues: linking positive selection signals and lineage diversification in the thymus. Immunol Rev 2006;209:253–273. [PubMed: 16448547]

5. Viola A, Contento RL, Molon B. T cells and their partners: The chemokine dating agency. Trends Immunol 2006;27:421–427. [PubMed: 16860609]

6. McHeyzer-Williams LJ, McHeyzer-Williams MG. Antigen-specific memory B cell development. Annu Rev Immunol 2005;23:487–513. [PubMed: 15771579]

7••. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, Maskell DJ, Simons RW, Cotter PA, Parkhill J, et al. Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science 2002;295:2091–2094. [PubMed: 11896279]This report describes the discovery of tropism switching *Bordetella* bacteriophage and diversity-generating retroelements. By sequencing various tropic phages, a region of variability within the *mtd* gene is identified. The authors further demonstrate that a downstream reverse transcriptase is required for diversity generation

8. Nikolich-Zugich J, Slifka MK, Messaoudi I. The many important facets of T-cell repertoire diversity. Nat Rev Immunol 2004;4:123–132. [PubMed: 15040585]

9. Williams CL, Boucher PE, Stibitz S, Cotter PA. BvgA functions as both an activator and a repressor to control Bvg phase expression of bipA in Bordetella pertussis. Mol Microbiol 2005;56:175–188. [PubMed: 15773988]

10. Diavatopoulos DA, Cummings CA, Schouls LM, Brinig MM, Relman DA, Mooi FR. Bordetella pertussis, the causative agent of whooping cough, evolved from a distinct, human-associated lineage of B. bronchiseptica. PLoS Pathog 2005;1:e45. [PubMed: 16389302]

11. Cummings CA, Bootsma HJ, Relman DA, Miller JF. Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. J Bacteriol 2006;188:1775–1785. [PubMed: 16484188]

12. Diavatopoulos DA, Hijnen M, Mooi FR. Adaptive evolution of the Bordetella autotransporter pertactin. J Evol Biol 2006;19:1931–1938. [PubMed: 17040390]

13. Poolman JT, Hallander HO. Acellular pertussis vaccines and the role of pertactin and fimbriae. Expert Rev Vaccines 2007;6:47–56. [PubMed: 17280478]

14. Hijnen M, Mooi FR, van Gageldonk PG, Hoogerhout P, King AJ, Berbers GA. Epitope structure of the Bordetella pertussis protein P. 69 pertactin, a major vaccine component and protective antigen. Infect Immun 2004;72:3716–3723. [PubMed: 15213111]

15. Mattoo S, Cherry JD. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to Bordetella pertussis and other Bordetella subspecies. Clin Microbiol Rev 2005;18:326–382. [PubMed: 15831828]

16. Liu M, Gingery M, Doulatov SR, Liu Y, Hodes A, Baker S, Davis P, Simmonds M, Churcher C, Mungall K, et al. Genomic and genetic analysis of Bordetella bacteriophages encoding reverse transcriptase-mediated tropism-switching cassettes. J Bacteriol 2004;186:1503–1517. [PubMed: 14973019]

17••. Doulatov S, Hodes A, Dai L, Mandhana N, Liu M, Deora R, Simons RW, Zimmerly S, Miller JF. Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. Nature 2004;431:476–481. [PubMed: 15386016]This report describes the role of the IMH sequence in specifying the directionality of template-mediated information transfer during diversity generation, and provides evidence that mutagenic homing results in the creation of mosaic VRs containing patches of diversified sequence. Furthermore, conserved DGR cassettes are identified in numerous phage and bacterial genomes

18. Kiss C, Fisher H, Pesavento E, Dai M, Valero R, Ovecka M, Nolan R, Phipps ML, Velappan N, Chasteen L, et al. Antibody binding loop insertions as diversity elements. Nucleic Acids Res 2006;34:e132. [PubMed: 17023486]

19••. McMahon SA, Miller JL, Lawton JA, Kerkow DE, Hodes A, Marti-Renom MA, Doulatov S, Narayanan E, Sali A, Miller JF, et al. The C-type lectin fold as an evolutionary solution for massive sequence variation. Nat Struct Mol Biol 2005;12:886–892. [PubMed: 16170324]The authors crystallized and present structures of five variant Mtd proteins including the pertactin-binding receptor from *Bordetella* bacteriophage BPP-1. This is the first characterization of the structures of DGR-diversified proteins and establishes the Clec domain as a stable scaffold to display variable residues

20. Robinson MJ, Sancho D, Slack EC, LeibundGut-Landmann S, Reis e Sousa C. Myeloid C-type lectins in innate immunity. Nat Immunol 2006;7:1258–1265. [PubMed: 17110942]

21. Vasta GR, Ahmed H, Odom EW. Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates. Curr Opin Struct Biol 2004;14:617–630. [PubMed: 15465324]

22•. Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. Febs J 2005;272:6179–6217. [PubMed: 16336259]This is a beautiful review covering the identification, structure, functional diversity, chemical dependence, ligand recognition repertoire, and classification of Clec domains within various protein families

23. Setubal JC, Reis M, Matsunaga J, Haake DA. Lipoprotein computational prediction in spirochaetal genomes. Microbiology 2006;152:113–121. [PubMed: 16385121]

24. Cullen PA, Haake DA, Adler B. Outer membrane proteins of pathogenic spirochetes. FEMS Microbiol Rev 2004;28:291–318. [PubMed: 15449605]

25. Han JS, Boeke JD. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? Bioessays 2005;27:775–784. [PubMed: 16015595]

26•. Lambowitz AM, Zimmerly S. Mobile group II introns. Annu Rev Genet 2004;38:1–35. [PubMed: 15568970]This comprehensive review summarizes molecular mechanisms responsible for group II intron mobility and includes comparisons with the mobility mechanisms utilized by retrotransposons and eukaryotic spliceosomal introns

27. Weiss RA. The discovery of endogenous retroviruses. Retrovirology 2006;3:67. [PubMed: 17018135]

28. Hasler J, Strub K. Alu elements as regulators of gene expression. Nucleic Acids Res 2006;34:5491–5497. [PubMed: 17020921]

29. Volff JN. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 2006;28:913–922. [PubMed: 16937363]

30. Le Rouzic A, Capy P. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. Genetics 2005;169:1033–1043. [PubMed: 15731520]

31. Black LW, Peng G. Mechanistic coupling of bacteriophage T4 DNA packaging to components of the replication-dependent late transcription machinery. J Biol Chem 2006;281:25635–25643. [PubMed: 16807240]

32. Martinez-Abarca F, Barrientos-Duran A, Fernandez-Lopez M, Toro N. The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. Nucleic Acids Res 2004;32:2880–2888. [PubMed: 15155857]

33. Christensen SM, Ye J, Eickbush TH. RNA from the 5′ end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. Proc Natl Acad Sci U S A 2006;103:17602–17607. [PubMed: 17105809]

34. Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 2006;357:1383–1393. [PubMed: 16490214]

35. Mieczkowski PA, Lemoine FJ, Petes TD. Recombination between retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae. DNA Repair (Amst) 2006;5:1010–1020. [PubMed: 16798113]

36. Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. RNA-templated DNA repair. Nature. 2007

37. Bernstein DA, Keck JL. Conferring substrate specificity to DNA helicases: role of the RecQ HRDC domain. Structure 2005;13:1173–1182. [PubMed: 16084389]

38. Weiss KK, Chen R, Skasko M, Reynolds HM, Lee K, Bambara RA, Mansky LM, Kim B. A role for dNTP binding of human immunodeficiency virus type 1 reverse transcriptase in viral mutagenesis. Biochemistry 2004;43:4490–4500. [PubMed: 15078095]

39. Boyer PL, Sarafianos SG, Clark PK, Arnold E, Hughes SH. Why do HIV-1 and HIV-2 use different pathways to develop AZT resistance? PLoS Pathog 2006;2:e10. [PubMed: 16485036]

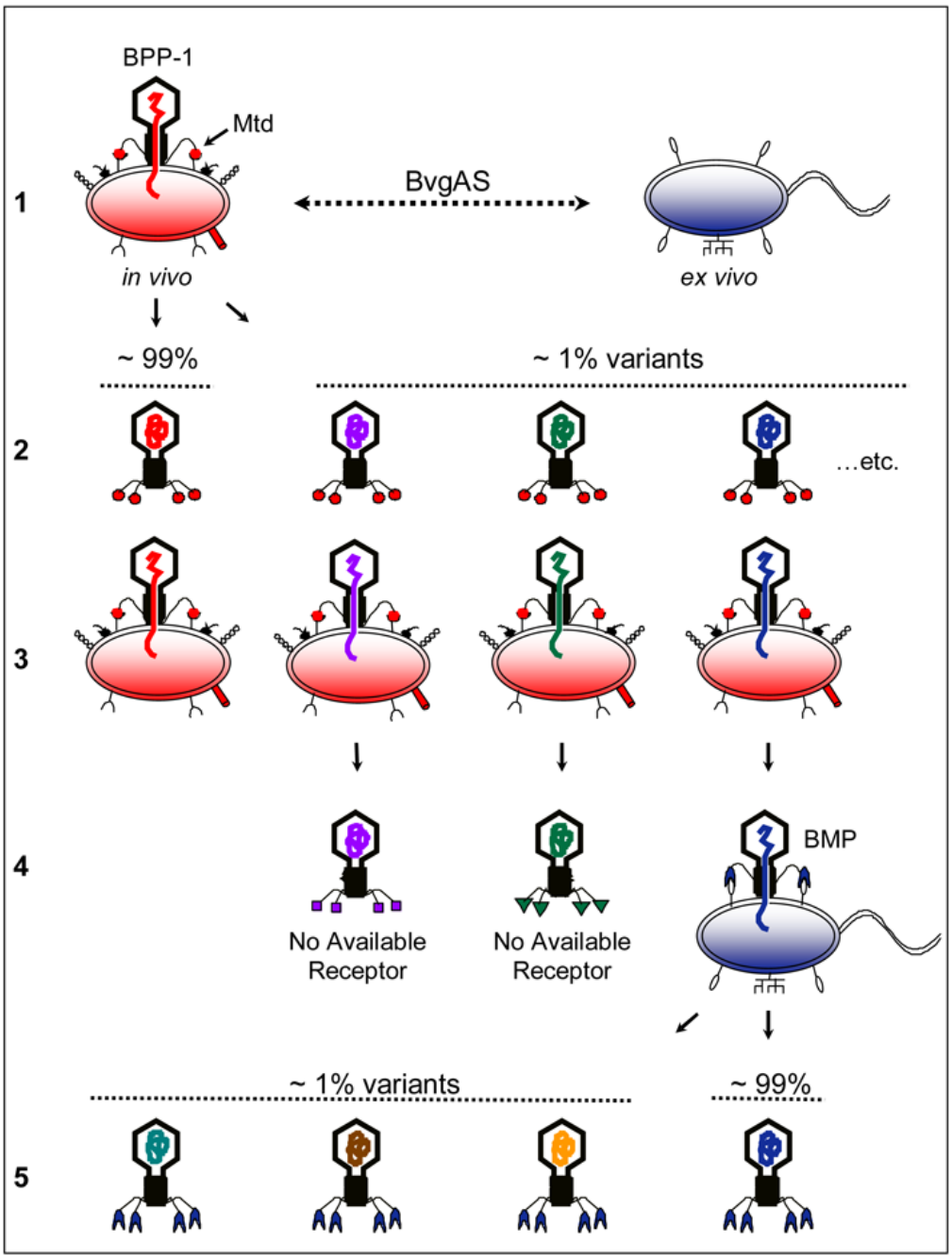

Figure 1.

DGR mediated tropism switching by *Bordetella* bacteriophage. **(1)** The BPP-1 Mtd tail fiber protein binds to pertactin on the surface of Bvg$^+$ phase *Bordetella* with subsequent injection of phage DNA. **(2)** Following phage genome replication and DGR function, ~1% of progeny phage contain a variant *mtd* (colored DNA). Since the frequency of diversification is low, parental genomes are expected to be in the vast majority in phage-producing cells, and variant genomes are likely to be packaged in virions containing the parental Mtd [31]. **(3)** Irrespective of genotype, the tropism specificity of the parent phage is retained for the second round of infection. **(4)** In the second round of infection, genomes diversified in the first round are finally packaged into virions with the variant Mtd molecules they encode. Approximately 1 in

1,000,000 of these progeny will express a novel Mtd that recognizes a receptor expressed on the surface of Bvg- phase *Bordetella*. **(5)** Ensuing phage infection and replication cycles will continue to generate Mtd variants.

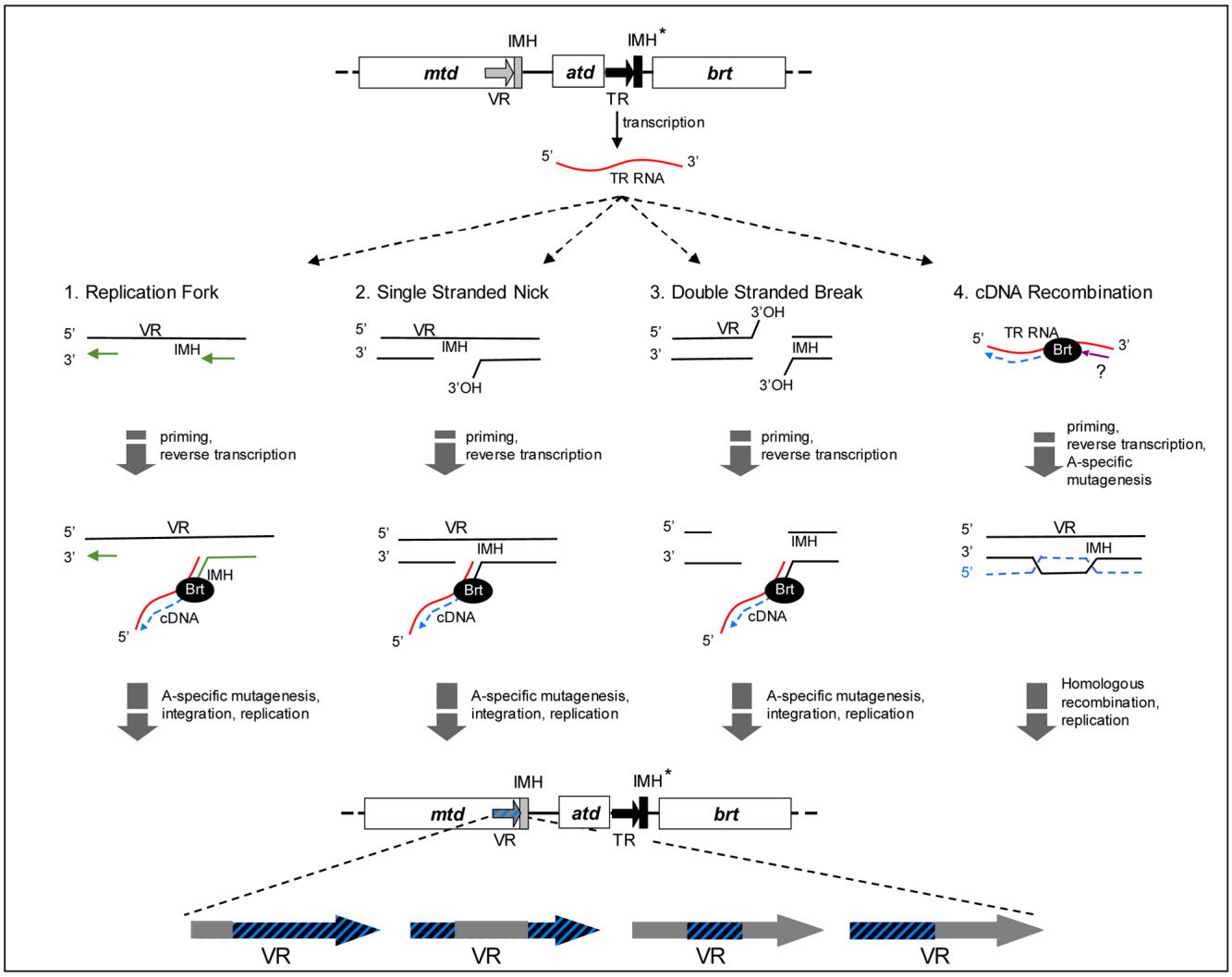

Figure 2.

Potential mechanisms of DGR retrotransposition. Four potential mechanisms for cDNA priming and mutagenic homing are shown. **(1)** Replication Fork. An Okazaki fragment (green arrow) serves as the initiating primer as the TR RNA transcript anneals to the recently synthesized antisense IMH complementary strand [32]. **(2)** Single Stranded Nick. A nick in the antisense strand occurs at IMH by a single-strand endonuclease that has yet to be identified. The resulting 3′ hydroxyl serves as the initiating primer. **(3)** Double Stranded Break. A yet to be determined factor creates a double strand break at the IMH. Progressive exonuclease degradation of the sense strand is accompanied by TR RNA transcript priming by the free 3′ hydroxyl on the antisense strand [33,34]. **(4)** cDNA Recombination. Brt reverse transcription of the TR RNA transcript is initiated by an unknown primer. VR variants are created by RecA independent homologous recombination of TR derived cDNA with the parental VR [35,36]. In all cases, DGR homing is followed by replication that produces mosaic VRs with patches of TR-derived variable sequence [••17].

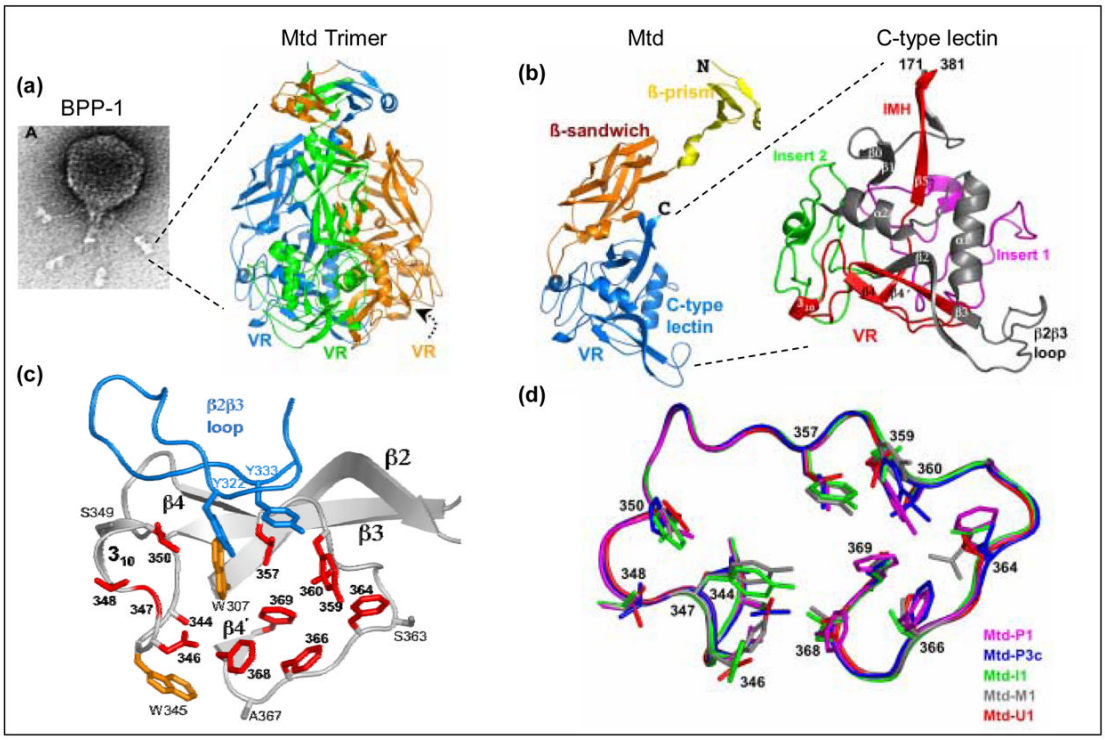

Figure 3.

Structure of Mtd [••19]. **(a)** Left: BPP-1 with bi-lobed globular structures at the distal ends of each tail fiber; Right: Each globular structure corresponds to a single Mtd trimer with three VR regions present on the bottom face. **(b)** Left: Mtd monomer; Right: Mtd CLec domain, VR sequences in red. The β5 strand, located at the very C-terminus of Mtd, is encoded by the 21 bp IMH sequence that sets the directionality of information transfer. Positioned in the central core of the trimer, β5 makes close intra- and inter-molecular contacts that would be disrupted by variation and, despite having adenine-encoded amino acids, IMH remains invariant. **(c)** Ligand-binding surface of an Mtd variant that binds pertactin. Variable side chains (red) are solvent exposed. Although the Mtd variant shown has a preponderance of hydrophobic variable residues, highly hydrophilic binding sites can also be generated. **(d)** Superimposed VR regions from five Mtd variants that bind different ligands. The main chain conformation of the Clec domain is remarkably invariant despite large differences in the chemical nature of variant side chains.

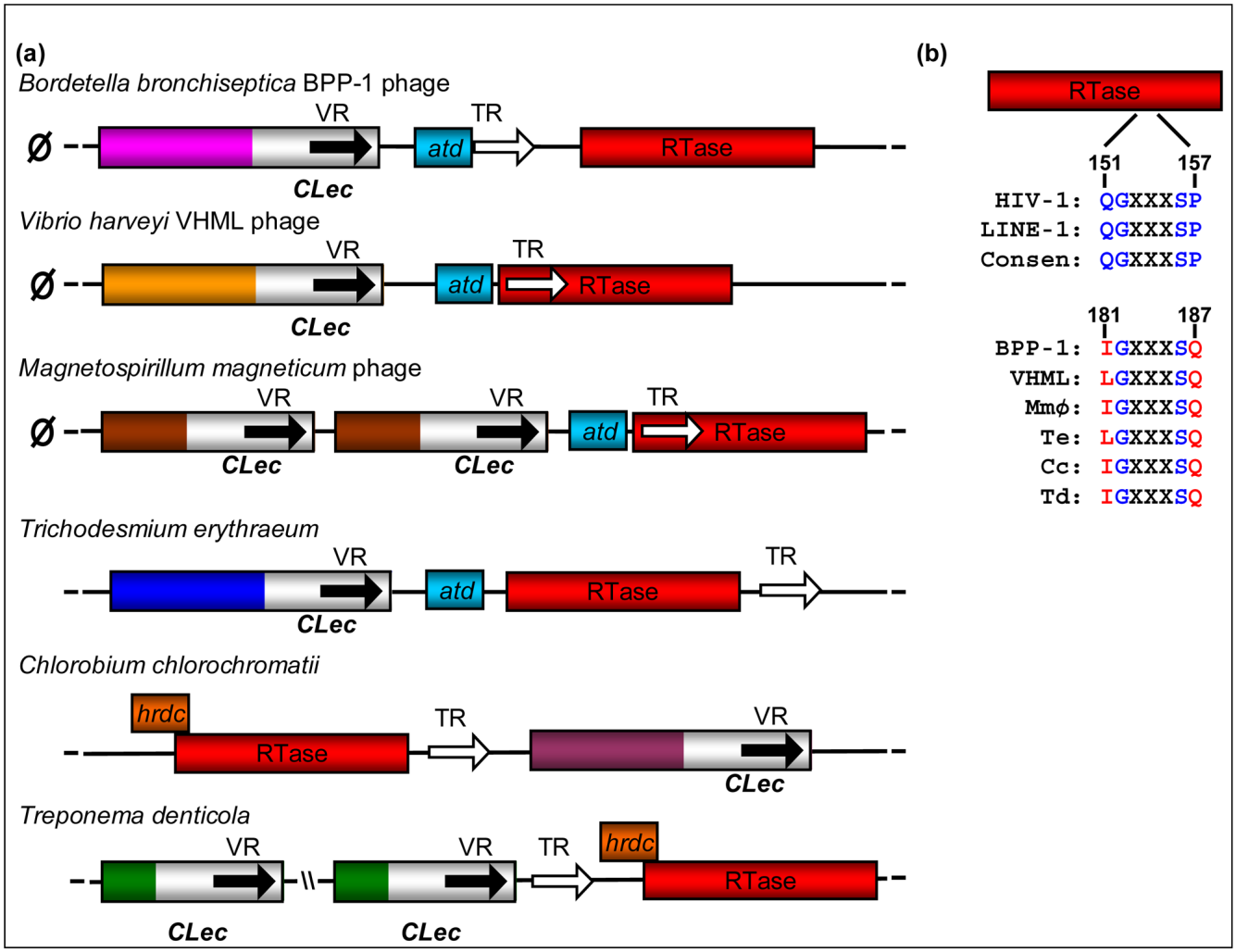

Figure 4.

Conserved and variable features of DGRs (*Xu M et al.*, unpublished data, *Gingery M et al.*, unpublished data) [••17]. **(a)** The prototype BPP-1 DGR is shown at the top. Orthologous reverse transcriptases (RTase) are located near predicted VRs and TRs. VRs are predominantly located within CLec domains at the C-terminal end of VR-containing ORFs. VR and TR pairs differ almost exclusively at positions correlating to adenines within the TR. Atd homologs are more divergent than the reverse transcriptase genes and they are present in a subset of DGRs. *Hrdc* loci, predicted to encode proteins with 80 amino acid helicase and RNAseD C-terminal (HRDC) domains, overlap RTase loci in a subset of DGRs [37]. **(b)** The short sequence represented by Gln151-Pro157 is highly conserved in HIV-1, LINE-1 and other "canonical" RTs, playing a pivotal role in dNTP and template recognition [38]. In particular, Q151 is located in the dNTP binding pocket and it directly interacts with the sugar moieties of incoming dNTPs. Mutations at this site (e.g. Q151M) confer resistance to nucleoside inhibitors such as AZT and they alter both mutational specificity and fidelity [38,39]. Although other shared unique features are evident, it may be significant that DGR-associated RTases share nearly identical polymorphisms (Q151 to I/L181, P157 to Q187) in a region known to play such a major role in controlling RT fidelity.