Results and Conclusions

The Gypsy Database (GyDB) is a wiki-style database launched with the aim of classifying non-redundant viruses and mobile genetic elements (MGEs) on the basis of their phylogenetic profile. Owing to the diversity of these genetic agents, the GyDB is a long term project. In a previous article, the second release of this project (GyDB 2.0), which focuses on the evolutionary classification of retroviruses and retrotransposons with long terminal repeats (LTRs) and their viral and host gene relatives in distinct organisms, was introduced. GyDB 2.0 is concerned with studying mobile elements within genomes. Therefore, an in-progress repository was created for databases with annotations of mobile genetic elements from particular genomes. This repository is called Mobilomics and the first uploaded database contains 549 LTR retroelements and related transposases which have been annotated from the genome of the Pea aphid Acyrthosiphon pisum. Mobilomics is accessible from the GyDB 2.0 project using the URL: http://gydb.org/index.php/Mobilomics.

Mobilomics is a tool undergoing continuous progression, which now contains the genome of the Pea aphid Acyrthosiphon pisum from which 549 LTR retroelement-like features were mapped and annotated. Gene identifiers in this collection follow the nomenclature provided by the International Aphid Genomic Consortium (IAGC) and this nomenclature was adopted to also deposit (as members of the IACG) annotations from our group into Aphidbase version 1.2. The collection consists of full-length and fragmented elements. Almost all annotations correspond to LTR retroelement sequences classified as members of the
typical LTR retroelement regions but differing from other Ty3/Gypsy elements as they possess an additional pol-like domain between the protease (AP) and RT. BLASTp2 searches performed using the NCBI protein databases revealed similarity (9.06e-17) between this domain and the Macro (MAC) domain of the Poa1p family (classified as cd02901 in the NCBI Conserved Domain Databases5). MAC is and Micropia/Mdg3 (Fig. 2A). In addition, four new clades named ATYMP01, ATYMP02 (Aphid Ty3/Gypsy element with Macro domain, Poa1p-like family), ATY3 and ATY4 (Aphid Ty3/Gypsy element) were identified. Figure 2B presents a graphical representation of these new elements. Remarkably, the ATYMP01 and ATYMP02 clades contain LTR retrotransposons and retroviruses with typical LTR retroelement regions but differing from other Ty3/Gypsy elements as they possess an additional pol-like domain between the protease (AP) and RT. BLASTp2 searches performed using the NCBI protein databases revealed similarity (9.06e-17) between this domain and the Macro (MAC) domain of the Poa1p family (classified as cd02901 in the NCBI Conserved Domain Databases5). MAC is
a high-affinity ADP-ribose binding module present in eukaryotes, prokaryotes and diverse RNA viruses, such as coronaviruses and alphaviruses, which replicate in the cytoplasm of animal cells. LTR retroelement-sequences with a MAC domain have only been reported in diverse Danio rerio retroviruses, where MAC maps between the Integrase (INT) and the ENV domains. Therefore, as far as we are aware, this is the first time that insect Ty3/Gypsy elements have been reported to contain MAC. Regarding Bel/Pao LTR retroelements, 86 full-length elements (two being putative retroviruses) and 92 additional partial Bel/Pao-like features, were annotated. In accordance with references 8 and 9, the Bel/Pao family can be divided into five major lineages—Bel, Suzu, Sinbad, Pao and Tas. As demonstrated in Figure 2C, all Bel/Pao sequences that were analyzed (on the basis of RT) are related to the Bel lineage. However, almost all the Bel/Pao elements that were characterized...
were new sequences that could be divided into five clades that were named ABP1, -2, -3, -4 and -5 (Aphid Bel/Pao element).

Graphical depictions are presented in Figure 2D. These new clades suggest that the diversity of the Bel/Pao family in Metazoans is greater than previously thought (as argued by de la Chaux and Wagner in a recent publication in ref. 10). It was interesting that although partial sequences related to the Tyl1/Copia family were annotated, no full-length Tyl1/Copia elements were detected in the screenings. These sequences are likely to represent reminiscences of an ancient Tyl1/Copia colonization but could be the case that further releases of the A. pisum genome (or other aphid relatives) will reveal the presence of full-length Tyl1/Copia elements.

The aphid collection includes multiple distinct single gene features that are likely to be of interest to researchers investigating the contribution of MGEs to the complexity of their host genomes. A putative case study, based on a pool of 60 sequences, has been selected and this contains a variety of LTR retroelement Solo-INTs, diverse cut-and-paste DNA transposons of the Ginger1 and -2 families, and other sequences identified as the chromodomain-carrying transposases (TRs) typically encoded by the Maverick/Polinton transposons. The latter TRs were classified using the term CIN1 (chromodomain-INTs type 1) as no apparent full-length Maverick/Polinton element was detected in the first A. pisum release. CIN1 TRs are therefore putative remnants derived from an ancient Maverick colonization. Interestingly, it is now known that an evolutionary link exists among Ginger-like and CIN1 TRs and LTR retroelement INTs, but this is also the case for other DNA transposons and INT-like host genes that are probably derived from the domestication of MGEs (reviewed in ref. 11 and 14–18). All of these are DDE INTs and TRs that, on the basis of INT-like structural or sequence similarities, can be considered as representative members of the Retroviral Integrate Superfamily of nucleic acid-processing enzymes involved in selfish evolution, replication and repair of DNA, recombination and gene fusion, RNA-mediated gene silencing and oncogenesis. The alignment used to infer this tree has been deposited in the GyDB collection with the URL: http://gydb.org/index.php/Collection_alignments. Red edges correspond to A. pisum sequences and black edges refer to reference INTs and TRs.

Genomic continuum representation of the scaffold 2,047 of the first aphid assembly, where GIN7d and GIN6c (nucleotides 34,807–36,407 and 38,019–36,918, respectively) map close to other MGEs not related to LTR retroelements. (C) Genomic continuum representation of the scaffold 6092, where the Maverick/Polinton-like TRs CIN1d and CIN1f map (nucleotides 28,451–29,308 and 30,520–31,959 respectively) map close to other features.

These new clades suggest that the diversity of the Bel/Pao family in Metazoans is greater than previously thought (as argued by de la Chaux and Wagner in a recent publication in ref. 10). It was interesting that although partial sequences related to the Tyl1/Copia family were annotated, no full-length Tyl1/Copia elements were detected in the screenings. These sequences are likely to represent reminiscences of an ancient Tyl1/Copia colonization but could be the case that further releases of the A. pisum genome (or other aphid relatives) will reveal the presence of full-length Tyl1/Copia elements.

The aphid collection includes multiple distinct single gene features that are likely to be of interest to researchers investigating the contribution of MGEs to the complexity of their host genomes. A putative case study, based on a pool of 60 sequences, has been selected and this contains a variety of LTR retroelement Solo-INTs, diverse cut-and-paste DNA transposons of the Ginger1 and -2 families, and other sequences identified as the chromodomain-carrying transposases (TRs) typically encoded by the Maverick/Polinton transposons. The latter TRs were classified using the term CIN1 (chromodomain-INTs type 1) as no apparent full-length Maverick/Polinton element was detected in the first A. pisum release. CIN1 TRs are therefore putative remnants derived from an ancient Maverick colonization. Interestingly, it is now known that an evolutionary link exists among Ginger-like and CIN1 TRs and LTR retroelement INTs, but this is also the case for other DNA transposons and INT-like host genes that are probably derived from the domestication of MGEs (reviewed in ref. 11 and 14–18). All of these are DDE INTs and TRs that, on the basis of INT-like structural or sequence similarities, can be considered as representative members of the Retroviral Integrate Superfamily of nucleic acid-processing enzymes involved in selfish evolution, replication and repair of DNA, recombination and gene fusion, RNA-mediated gene silencing and oncogenesis. The alignment used to infer this tree has been deposited in the GyDB collection with the URL: http://gydb.org/index.php/Collection_alignments. Red edges correspond to A. pisum sequences and black edges refer to reference INTs and TRs.

Genomic continuum representation of the scaffold 2,047 of the first aphid assembly, where GIN7d and GIN6c (nucleotides 34,807–36,407 and 38,019–36,918, respectively) map close to other MGEs not related to LTR retroelements. (C) Genomic continuum representation of the scaffold 6092, where the Maverick/Polinton-like TRs CIN1d and CIN1f map (nucleotides 28,451–29,308 and 30,520–31,959 respectively) map close to other features.

Figure 2D. These new clades suggest that the diversity of the Bel/Pao family in Metazoans is greater than previously thought (as argued by de la Chaux and Wagner in a recent publication in ref. 10). It was interesting that although partial sequences related to the Tyl1/Copia family were annotated, no full-length Tyl1/Copia elements were detected in the screenings. These sequences are likely to represent reminiscences of an ancient Tyl1/Copia colonization but could be the case that further releases of the A. pisum genome (or other aphid relatives) will reveal the presence of full-length Tyl1/Copia elements.

The aphid collection includes multiple distinct single gene features that are likely to be of interest to researchers investigating the contribution of MGEs to the complexity of their host genomes. A putative case study, based on a pool of 60 sequences, has been selected and this contains a variety of LTR retroelement Solo-INTs, diverse cut-and-paste DNA transposons of the Ginger1 and -2 families, and other sequences identified as the chromodomain-carrying transposases (TRs) typically encoded by the Maverick/Polinton transposons. The latter TRs were classified using the term CIN1 (chromodomain-INTs type 1) as no apparent full-length Maverick/Polinton element was detected in the first A. pisum release. CIN1 TRs are therefore putative remnants derived from an ancient Maverick colonization. Interestingly, it is now known that an evolutionary link exists among Ginger-like and CIN1 TRs and LTR retroelement INTs, but this is also the case for other DNA transposons and INT-like host genes that are probably derived from the domestication of MGEs (reviewed in ref. 11 and 14–18). All of these are DDE INTs and TRs that, on the basis of INT-like structural or sequence similarities, can be considered as representative members of the Retroviral Integrate Superfamily of nucleic acid-processing enzymes involved in selfish evolution, replication and repair of DNA, recombination and gene fusion, RNA-mediated gene silencing and oncogenesis. The alignment used to infer this tree has been deposited in the GyDB collection with the URL: http://gydb.org/index.php/Collection_alignments. Red edges correspond to A. pisum sequences and black edges refer to reference INTs and TRs.

Genomic continuum representation of the scaffold 2,047 of the first aphid assembly, where GIN7d and GIN6c (nucleotides 34,807–36,407 and 38,019–36,918, respectively) map close to other MGEs not related to LTR retroelements. (C) Genomic continuum representation of the scaffold 6092, where the Maverick/Polinton-like TRs CIN1d and CIN1f map (nucleotides 28,451–29,308 and 30,520–31,959 respectively) map close to other features.

Figure 3 (See opposite page). (A) Inferred phylogenetic tree based on solo-INTs and TRs of A. pisum using identical methods to those described in Figure 2. The alignment used to infer this tree has been deposited in the GyDB collection with the URL: http://gydb.org/index.php/Collection_alignments. Red edges correspond to A. pisum sequences and black edges refer to reference INTs and TRs. (B) Genomic continuum representation of the scaffold 2,047 of the first aphid assembly, where GIN7d and GIN6c (nucleotides 34,807–36,407 and 38,019–36,918, respectively) map close to other MGEs not related to LTR retroelements. (C) Genomic continuum representation of the scaffold 6092, where the Maverick/Polinton-like TRs CIN1d and CIN1f map (nucleotides 28,451–29,308 and 30,520–31,959 respectively) map close to other features.
References


