Analytical Biochemistry 600 (2020) 113770

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Impact of transposable elements on genome size variation between two closely related crustacean species

Thomas Becking[a], Clément Gilbert[a,b], Richard Cordaux[a,*]

[a] Laboratoire Ecologie et Biologie des Interactions, UMR CNRS 7267, Université de Poitiers, Poitiers, France
[b] Present Address: Laboratoire Evolution, Génomes, Comportement, Écologie, Unité Mixte de Recherche 9191 Centre National de La Recherche Scientifique and Unité Mixte de Recherche 247 Institut de Recherche pour le Développement, Université Paris-Sud, 91198, Gif-sur-Yvette, France

ARTICLE INFO

Keywords:
Transposable elements
Genome sequencing
Comparative analysis
Crustaceans
Terrestrial isopods
Genome size

ABSTRACT

Identifying and quantifying genome size variation among species and understanding the underlying causes is a long-standing objective in evolutionary biology. Here, we investigated the basis of genome size variation between two closely related species of terrestrial isopods: *Armadillidium vulgare* and *Armadillidium nasatum*. The two species diverged 25 million years ago and the *A. vulgare* genome is ~500 megabases larger than the *A. nasatum* genome (1.7 vs. 1.2 gigabases, respectively). Our analyses indicated that genome size difference is essentially attributed to transposable elements (TEs). We found that the deletion rate may be slightly higher in *A. nasatum* than in *A. vulgare*, but it is unlikely to explain the observed genome size difference. As the two genomes largely share the same TE families, differential transpositional activity also contributes to the observed variation. Analyses of TE expression suggested that the cumulative expression level of all expressed TEs was higher in *A. nasatum* than in *A. vulgare*. Assuming TE expression level is a good proxy for TE transpositional activity, our results suggest that the two species may have recently been experiencing different TE transposition dynamics. Overall, our results illustrate the important impact TEs can have on genome structure and evolution between closely related species.

1. Introduction

Genome size variation is one of the most noteworthy sources of biodiversity on Earth [1]. The nature and importance of the evolutionary forces and the molecular mechanisms causing this variation are the subject of numerous studies and debates [2–8]. Many studies have shown that genome size is positively correlated to the content in transposable elements (TEs), as for example in mammals [9], birds [9], insects [10–16], or flowering plants [17,18]. However, the reasons explaining why TEs accumulate more in some species than in others remain largely unknown.

To date, more than 4000 eukaryote genomes have been sequenced and are available in GenBank (last accessed on December 2019). Crustaceans, however, are poorly represented among these genomes. Despite the existence of more than 70,000 living species of crustaceans [19], the genomes of only ~30 species are available to date. While crustaceans are primarily marine organisms, constituting a large proportion of ocean biomass [20], many species have colonized freshwater, semi-terrestrial or terrestrial environments. Recently, the genomes of two species of terrestrial isopods have been assembled: *Armadillidium vulgare* [21] and *Armadillidium nasatum* [22]. These two species are closely related and they diverged from a common ancestor ~25 million years ago [23]. However, it turns out that the genome assembly of *A. vulgare* (1.7 Gb) [21] is ~500 megabases larger than that of *A. nasatum* (1.2 Gb) [22]. In both cases, the genomes were sequenced using the same hybrid strategy combining Illumina short reads and PacBio long reads, and following the same assembly protocol [21,22]. These resources offer an interesting opportunity to compare genome size evolution between two closely related species. In this study, we used a comparative approach to investigate the causes of the large genome size variation observed between the two closely related species of terrestrial isopods *A. vulgare* and *A. nasatum*. Our results indicate that differential transpositional activity of the same TE families likely underlies most of the variation in genome size between the two species.

Abbreviations: piRNA, PIWI-interacting RNA; RPKM, reads per kilobase of exon model per million mapped; TE, transposable element; TPM, transcripts per million

* Corresponding author. Université de Poitiers, Laboratoire Ecologie et Biologie des Interactions - UMR CNRS 7267, Equipe Ecologie Evolution Symbiose, Bâtiment B8-B35, 5 Rue Albert Turpain, TSA 51106, 86073, Poitiers Cedex 9, France.
E-mail address: richard.cordaux@univ-poitiers.fr (R. Cordaux).

https://doi.org/10.1016/j.ab.2020.113770
Received 23 December 2019; Received in revised form 13 April 2020; Accepted 5 May 2020
Available online 08 May 2020

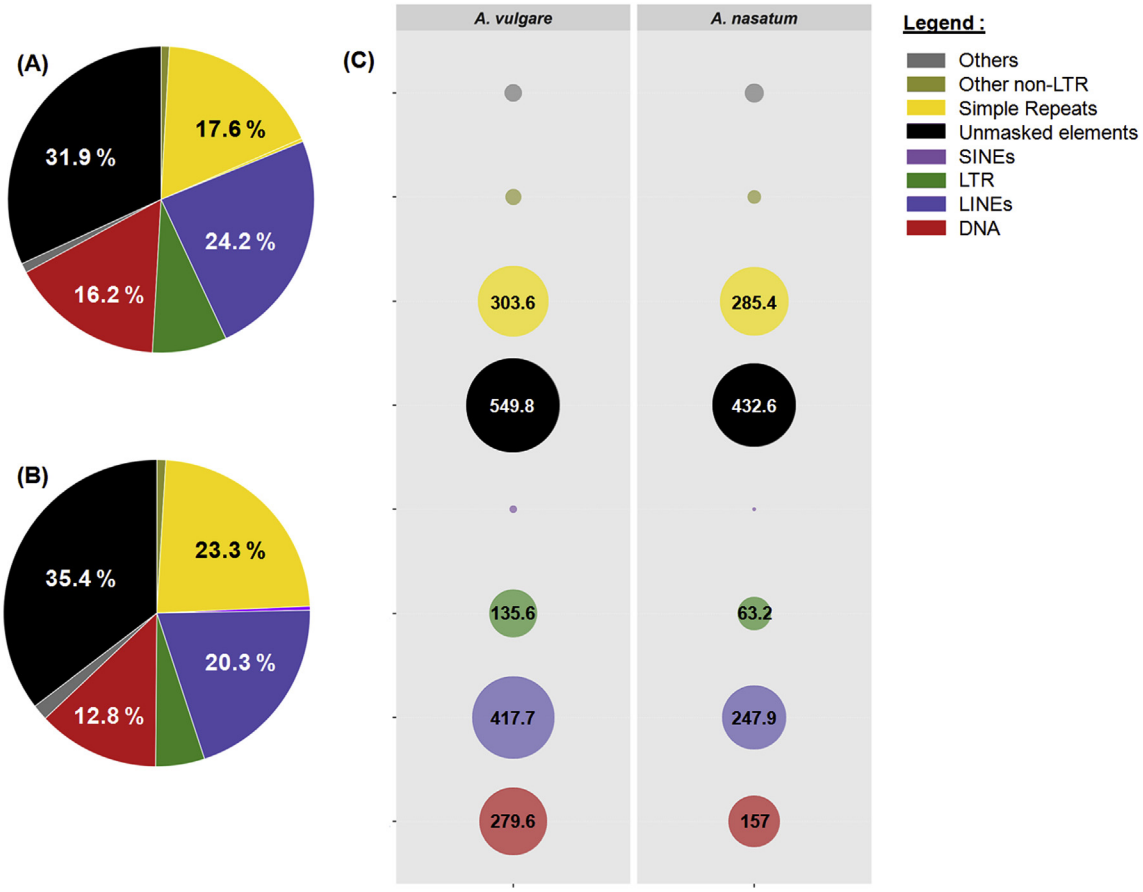

Fig. 1. Content of the *Armadillidium vulgare* (A) and *Armadillidium nasatum* (B) genomes. (C) Comparison of the cumulative length (in megabases) of various genomic categories in *A. vulgare* and *A. nasatum*.

2. Materials and methods

2.1. Genomic data

The genomes of *A. vulgare* (accession number: SAUD00000000) and *A. nasatum* (accession number: SEYY00000000) were both assembled using a hybrid approach combining short paired-end Illumina reads and long PacBio reads and the exact same suite of software, as described previously [21,22]. The *A. vulgare* assembly is composed of 43,541 contigs/scaffolds, totaling 1.725 Gb with an N_{50} (minimum contig length to cover 50% of the genome) of 51,088 bp. Based on a BUSCO analysis, we found that it includes 92.0% of the 1066 specific arthropod core genes [24]. The *A. nasatum* assembly is composed of 25,196 contigs/scaffolds, totaling 1.223 Gb with an N_{50} of 86,284 bp and including 93.9% of the 1066 specific arthropod core genes. Analyses of gene and repeat contents were based on the structural and functional annotations of both assemblies.

2.2. Analyses of TE expression

To identify the genomic origin of expressed TEs, we used the pipeline TEcandidates [25]. In brief, TEcandidates uses *de novo* transcriptome assembly to assess expressed TE instances. To restrict reads to repeat zones, TEcandidates first runs a pre-alignment of the reads on the genome and keeps the reads that map to RepeatMasker predictions. Then, a *de novo* transcriptome assembly is performed using Trinity v2.4.0 with default parameters [26]. Generated contigs are mapped to the reference genome, and their intersection with annotated TEs is assessed with BEDtools v2.26 [27]. We used the multifasta files corresponding to *A. vulgare* and *A. nasatum* genomes, the annotation gff files generated by RepeatMasker (as each genome annotation has resolution at the level of TE instance), and RNAseq datasets (100 bp-long single end reads) available from the literature [28], which were generated from whole bodies of females originating from various locations: 10 datasets for *A. vulgare* (SRA accession numbers: SRX564995 to SRX565004) and two for *A. nasatum* (SRX564993 and SRX564994). TEcandidates pipeline was then used with default parameters, as recommended by the authors instructions. RNAseq read quality was analyzed with FastQC (version 0.11.4, http://www.bioinformatics.babraham.ac.uk/projects/89fastqc). Removal of low-quality reads and sequencing adaptors was performed with Trimmomatic (version 0.33), setting the ILLUMINACLIP palindrome clip threshold at 30 and the simple clip threshold at 10. The quality threshold was set to a minimum Phred score of 20.

To quantify TE expression, we mapped the RNAseq reads onto the potentially expressed TEs identified by TEcandidates. Reads from multiple RNAseq datasets were mapped independently using Bowtie 2 [29] using the *very sensitive* mode (parameter: *end-to-end*). TE genomic coverage was calculated using BEDTools [27], and the Artemis software (v18.0.0) [30] was used to quantify expressed transcripts, and yielded reads per kilobase of exon model per million mapped (RPKM) expression values. Transcripts per million (TPM) were then calculated as $TPM_{(i)} = (RPKM_{(i)}/\Sigma RPKM)x10^6$.

3. Results & discussion

3.1. Major contribution of transposable elements to genome size variation

To investigate the possible factors underlying the difference in genome size between *A. vulgare* and *A. nasatum*, we first compared the

Table 1
Comparison of repeat content in the *Armadillidium vulgare* and *Armadillidium nasatum* genomes.

	Armadillidium vulgare (a)		*Armadillidium nasatum* (b)		Genome size difference (a-b)	
	Count	Cumulative length (bp)	Count	Cumulative length (bp)	Cumulative length (bp)	% cumulative length
Genome	n/a	1,725,108,002	n/a	1,223,175,971	501,932,031	n/a
All TEs	2,132,032	854,113,876	1,544,121	484,727,234	369,386,642	73.59%
DNA transposons	986,943	279,615,254	750,157	157,019,308	122,595,946	24.42%
LINEs	796,433	417,726,622	532,873	247,894,416	169,832,206	33.84%
SINEs	37,622	5,933,232	32,834	4,978,843	954,389	0.19%
LTRs	246,423	135,610,359	179,870	63,190,568	72,419,791	14.43%
Retroposons	64,611	15,228,409	48,387	11,644,099	3,584,310	0.71%
Tandem repeats	2,536,308	303,620,084	2,047,530	285,430,073	18,190,011	3.62%
Other repeats	227,683	17,582,771	217,427	20,436,134	−2,853,363	−0.57%

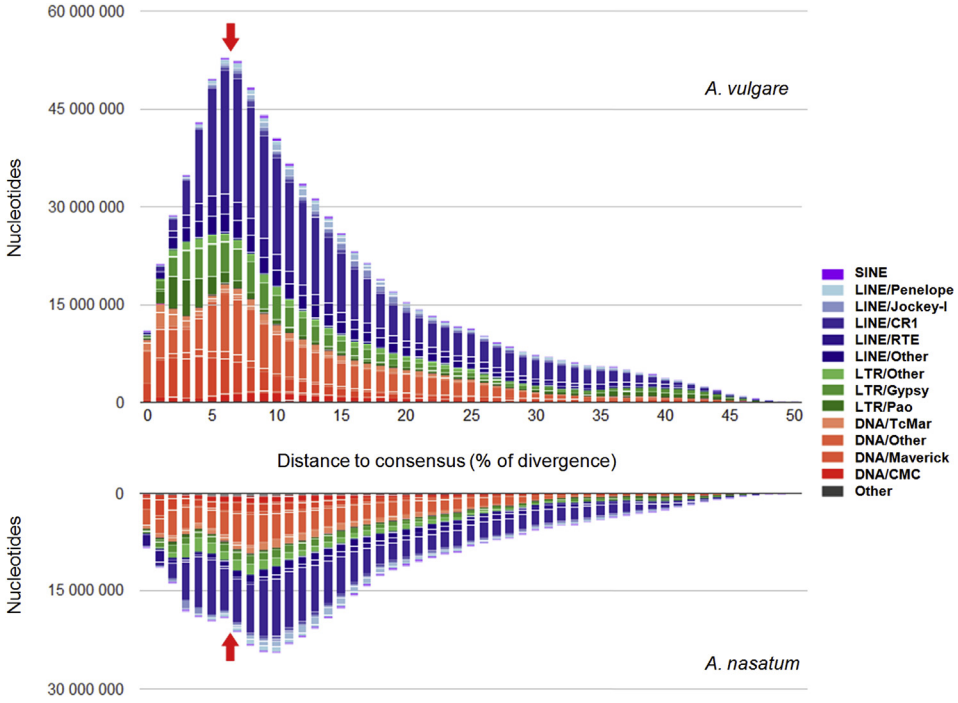

Fig. 2. Frequency distribution of transposable element families according to the divergence of individual copies to their respective family consensus sequences. Transposable elements identified in *Armadillidium vulgare* are shown in the upper part of the figure and those identified in *Armadillidium nasatum* in the lower part. The red arrow indicates the estimated genetic distance between *A. vulgare* and *A. nasatum*. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

gene content of the two species. The structural annotation identified 19,051 and 14,636 genes in *A. vulgare* and *A. nasatum*, respectively. The difference in gene number between the two genomes corresponds to 26.6 Mb, as all genes combined represent 164.5 Mb in *A. vulgare* [21] and 137.9 Mb in *A. nasatum* [22]. Thus, variation in gene number only accounts for 5% of the genome size difference between the two species. Comparison of tandem repeat content between *A. vulgare* and *A. nasatum* indicated similarly high amounts: 303.6 Mb (or 17.6% of the genome) for *A. vulgare* and 285.4 Mb (or 23.3% of the genome) for *A. nasatum* (Fig. 1, Table 1). Thus, tandem repeats do not contribute much to the difference in genome size between *A. vulgare* and *A. nasatum*. In fact, the difference in size between the two genomes is mainly explained by a much larger number of TE copies (~0.6 million) in *A. vulgare* relative to *A. nasatum*. Indeed, there are ~2.1 million TE copies (or fragments) annotated by RepeatMasker in *A. vulgare* vs. only ~1.5 million in *A. nasatum* (Table 1). Consequently, the genome of *A. vulgare* has 854.1 Mb derived from TEs while the genome of *A. nasatum* has only 484.7 Mb (Figs. 1 and 2, Table 1). Thus, 74% of the observed size difference between the two genomes is accounted for by TEs (Table 1).

Interestingly, the activity peak of TEs in *A. nasatum* corresponds to copies of TEs with ~10% divergence from their consensus sequence, while the activity peak of TEs in *A. vulgare*, which is much higher than that of *A. nasatum* TEs, corresponds to copies of TEs with ~6% divergence from their consensus sequence (Fig. 2). The age profile thus suggests that, unlike *A. nasatum*, *A. vulgare* underwent strong transpositional activity at about the time of, or not long after, the divergence between the two species (estimated to ~25 million years ago or ~6% divergence, see Ref. [23]). The higher amount of TEs observed in *A. vulgare* compared to *A. nasatum* could result from this transposition burst. Alternatively, *A. nasatum* and *A. vulgare* could have undergone similar TE activity through time but *A. nasatum* may have a higher deletion rate compared to *A. vulgare*. To compare deletion rates in the two genomes, we analyzed the size of all TE copies > 500 bp in length in *A. nasatum* (n = 321,878 copies) and *A. vulgare* (n = 546,470 copies), enabled by the fact that they share the same TE families. We found that the median values of TE copy sizes were similar between the two genomes (823 bp in *A. nasatum* and 868 bp in *A. vulgare*) but the difference was statistically significant (Mann-Whitney *U* test, $W = 4.78 \times 10^{10}$, $p < 10^{-16}$). These results suggested that the deletion rate of *A. nasatum* may be higher than that of *A. vulgare* and that it could contribute to the genome size difference between the two species. However, the difference in deletion rates between *A. nasatum* and *A. vulgare* is apparently modest. Therefore, it is unlikely to account for the whole variation in genome size observed between the two

Table 2
Comparison of transposable element (TE) copy numbers in the *Armadillidium vulgare* and *Armadillidium nasatum* genomes, for the 61 TE families with at least 1000 copies in both species.

TE family	*Armadillidium vulgare*		*Armadillidium nasatum*		*A. vulgare* to *A. nasatum* ratio of proportions
	Copy number	Proportion of all TE copies (%)	Copy number	Proportion of all TE copies (%)	
CR1-Zenon	337,294	22.81	235,598	20.59	1.11
Gypsy	115,386	7.80	83,358	7.28	1.07
hAT-Ac	115,117	7.78	72,005	6.29	1.24
CMC-EnSpm	95,687	6.47	84,698	7.40	0.87
RTE-BovB	72,051	4.87	34,177	2.99	1.63
Retroposon	54,611	3.69	48,387	4.23	0.87
Penelope	43,236	2.92	41,185	3.60	0.81
R2	42,048	2.84	41,099	3.59	0.79
Unknown	31,573	2.13	50,188	4.39	0.49
Maverick	31,231	2.11	31,812	2.78	0.76
Sola	31,007	2.10	9341	0.82	2.57
hAT	30,107	2.04	12,581	1.10	1.85
TcMar-Tc1	27,507	1.86	35,860	3.13	0.59
L2	27,085	1.83	29,429	2.57	0.71
Pao	25,502	1.72	5885	0.51	3.35
hAT-Charlie	23,954	1.62	18,734	1.64	0.99
Ginger	22,578	1.53	25,299	2.21	0.69
hAT-Tip 100	20,555	1.39	15,866	1.39	1.00
CR1	18,910	1.28	14,386	1.26	1.02
Merlin	18,711	1.27	8968	0.78	1.61
TcMar-Fot 1	16,186	1.09	14,061	1.23	0.89
hAT-hATm	15,675	1.06	11,660	1.02	1.04
Helitron	13,708	0.93	16,396	1.43	0.65
P	13,638	0.92	12,783	1.12	0.83
PIF-Harbinger	12,688	0.86	14,070	1.23	0.70
non-LTR	11,538	0.78	1376	0.12	6.49
CMC-Chapaev-3	10,973	0.74	9362	0.82	0.91
ERV1	10,825	0.73	11,098	0.97	0.75
CMC-Transib	10,685	0.72	9367	0.82	0.88
MULE-MuDR	10,496	0.71	11,553	1.01	0.70
L1	10,039	0.68	8028	0.70	0.97
TcMar	9587	0.65	2005	0.18	3.70
Jockey	9211	0.62	7642	0.67	0.93
Copia	8870	0.60	4017	0.35	1.71
hAT-Tol 2	8419	0.57	5125	0.45	1.27
CMC-Chapaev	7385	0.50	7094	0.62	0.81
hAT-Tag 1	7037	0.48	2628	0.23	2.07
ID	6017	0.41	1070	0.09	4.35
TcMar-Mariner	5933	0.40	6229	0.54	0.74
tRNA-Core	5280	0.36	5280	0.46	0.77
TcMar-Tigger	4984	0.34	3325	0.29	1.16
Novosib	4934	0.33	3011	0.26	1.27
hAT-Blackjack	4824	0.33	3493	0.31	1.07
ERVK	4544	0.31	5012	0.44	0.70
Dong-R4	4174	0.28	4032	0.35	0.80
I	4105	0.28	4253	0.37	0.75
DRE	3928	0.27	2483	0.22	1.22
Kolobok-Hydra	3853	0.26	4000	0.35	0.75
ERV	3846	0.26	1334	0.12	2.23
L1-Tx1	3508	0.24	4889	0.43	0.56
Dada	3021	0.20	3218	0.28	0.73
Kolobok-T2	3006	0.20	3377	0.30	0.69
tRNA	2713	0.18	6416	0.56	0.33
TcMar-m44	2128	0.14	1984	0.17	0.83
PIF-ISL2EU	1739	0.12	1708	0.15	0.79
IS3EU	1641	0.11	1643	0.14	0.77
DNA_virus	1621	0.11	2088	0.18	0.60
Proto 1	1464	0.10	2063	0.18	0.55
Crypton-V	1348	0.09	1095	0.10	0.95
R1	1241	0.08	1155	0.10	0.83
Ngaro	1020	0.07	1011	0.09	0.78

species. Thus, higher transpositional activity in *A. vulgare* than in *A. nasatum* may have also contributed to the observed genome size difference.

According to our analyses, the most represented TEs in both species belong to the LINE category (accounting for > 20% of each genome, Fig. 1). These LINEs are divided into 14 families shared by both species (Fig. 2). The second most represented TEs are DNA transposons, which represent 16.2% of the *A. vulgare* genome and 12.8% of the *A. nasatum* genome (Fig. 1). These DNA transposons belong to 17 families shared by both genomes (Fig. 2). The third most represented TEs are LTR retrotransposons which represent 5–8% of both genomes (Fig. 1), distributed in 10 families shared by both species (Fig. 2). In sum, the difference in TE numbers between *A. vulgare* and *A. nasatum* is not due to the acquisition and expansion of new TE families in *A. vulgare* relative to *A. nasatum*. To investigate whether the higher TE content in *A. vulgare* is due to the success of specific TE families, we focussed on the 61 TE families with at least 1000 copies in both *A. vulgare* and *A. nasatum* (Table 2). First, we calculated the proportion of each TE family relative to all TEs, in each genome. Then, we calculated the ratios of proportions of each TE family in *A. vulgare* relative to *A. nasatum*. TE families should have a ratio of one in the absence of enrichment in any species, > 1 if enriched in *A. vulgare* and < 1 if enriched in *A. nasatum*. We found that the median value of *A. vulgare* to *A. nasatum* ratios for the 61 TE families was very close to one (1.02). Specifically, only two TE families had ratios < 0.5 (suggesting substantial enrichment in *A. nasatum*) and only four TE families had ratios > 3 (suggesting substantial enrichment in *A. vulgare*). These 6 TE families combined accounted for < 6% of all TE copies in both genomes. We conclude that the success of specific TE families is unlikely to explain the higher TE content in *A. vulgare* relative to *A. nasatum*. Instead, differential amplification of the same TE families appears to underlie most variation in TE content between the two species. These results point to a difference in TE activity between the two genomes that would hold for all or most TE families. In arthropods, including isopod crustaceans, TE silencing in the germline occurs mainly through the action of PIWI-interacting RNAs (piRNAs) [31–33]. Ultimately, the observed differences in TE content between *A. nasatum* and *A. vulgare* could be due to variation in the strength of the piRNA pathway between the two species. Whether such variation is in favor of a global TE de-repression in *A. vulgare* or a stronger TE repression in *A. nasatum* is an interesting question that will be worth exploring in the future.

In line with a growing body of literature derived from diverse taxonomic groups ranging from vertebrates [9,34,35] to invertebrates [11–14,36] and plants [37–39], our analysis shows that genome size variation across two closely related crustacean species can largely be explained by differences in TE content. As the divergence time between *A. vulgare* and *A. nasatum* is estimated to ~25 million years and genome size difference due to TE abundance is roughly 370 Mb (Table 1), genome size divergence between *A. vulgare* and *A. nasatum* is estimated to ~15 Mb per million years. Considering genome sizes in the gigabase range, this rate may not seem stunning, but it appears to be strong enough to generate substantial genome size differences at an evolutionary time scale.

3.2. Comparison of TE expression profiles in A. vulgare and A. nasatum

While peaks of transpositional activity in *A. vulgare* and *A. nasatum* were inferred at 5–10% divergence from consensus sequences, the occurrence of TEs with much lower divergence levels suggests that transpositional activity may be ongoing (Fig. 2). Consistently, cases of TE horizontal transfer have been reported during the recent evolutionary history of terrestrial isopods [40]. The higher amount of TE copies at low divergence values in *A. vulgare* than in *A. nasatum* suggests higher recent transpositional activity in the former than in the latter (Fig. 2). If so, it may be predicted that TE expression levels may be higher in *A. vulgare* than in *A. nasatum*. To explore this possibility, we analyzed TE expression levels in both species. The TEcandidate pipeline identified 5317 TE copies (distributed in 4055 different contigs) potentially transcribed in *A. vulgare* (0.25% of the total number of TE copies) compared to 1259 in *A. nasatum* (0.08% of the total number of TE copies, distributed in 1101 genomic contigs) (Table 3). This analysis

Table 3
Comparison of expressed copies of transposable elements in the *Armadillidium vulgare* and *Armadillidium nasatum* genomes.

	Armadillidium vulgare		*Armadillidium nasatum*	
	Count	%	Count	%
LINEs	2822	53.08%	657	52.18%
SINEs	6	0.11%	0	0%
LTRs	779	14.65%	258	20.49%
DNA transposons	1374	25.84%	318	25.26%
Retroposons	314	5.91%	11	0.87%
Others	22	0.41%	15	1.19%
All transposable elements	5317	100%	1259	100%

suggested that there may be ~4 times more transcribed TE copies in *A. vulgare* than in *A. nasatum*. In both species, LINEs were the TEs with the highest number of transcribed TE copies, comprising > 50% of all transcribed copies (Table 3). In comparison, DNA transposons accounted for ~25% of all transcribed copies (Table 3).

To quantify TE expression levels in the two genomes, we carried out RPKM and TPM analyses based on RNAseq data. It appeared that the most highly expressed TEs in *A. vulgare* and *A. nasatum* are different. Indeed, CR1-Zenon LINEs are by far the most highly expressed TEs in *A. vulgare*, followed by unclassified DNA elements, unclassified LINEs, RTE-BovB LINEs and Gypsy LTR retrotransposons (Fig. 3). In contrast, unclassified DNA transposons are by far the most highly expressed TEs in *A. nasatum*, followed by CR1-Zenon LINEs, Gypsy LTR retrotransposons RTE-BovB LINEs and unclassified LINEs (Fig. 3). Interestingly, the cumulative expression level of all expressed TEs was higher in *A. nasatum* than in *A. vulgare*, with total cumulative RPKM values of 778,709 and 686,967, respectively. Thus, our results suggest that the species with the lowest, not the highest, TE amplification rate is associated with the current highest TE expression level. Assuming that TE expression level is a good proxy for TE transpositional activity, our results may reflect that TE transposition rate is varying in opposite directions in the two species: it may be declining in *A. vulgare* after a phase of intense amplification while it may be rising or steady in *A. nasatum*. As mentioned above, such variation could ultimately be caused by variation in the strength of the piRNA-mediated TE silencing pathway. It should be noted, however, that the RNAseq data we used were derived from female whole bodies [28]. Therefore, the recorded expression levels may not reflect solely germline expression but a mix of germline and somatic TE expression. Yet, gonads represent the major tissue in terrestrial isopod females, suggesting that the observed TE expression levels plausibly reflect germline TE expression to a large extent.

4. Conclusion

In summary, our analyses demonstrated that even though the *A.*

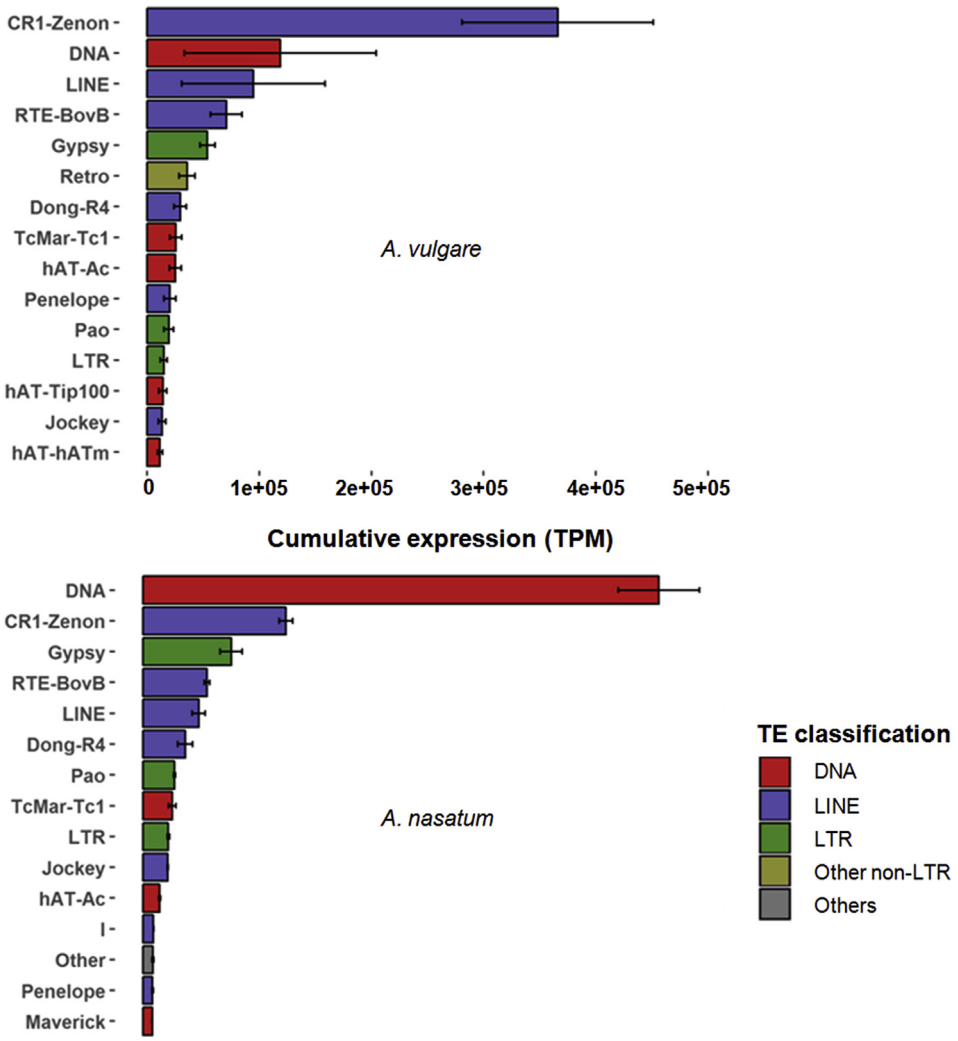

Fig. 3. Comparison of the most expressed transposable element families in the *Armadillidium vulgare* (top) and *Armadillidium nasatum* (bottom) genomes. The cumulative expression (in transcripts per million, TPM) of the 15 most expressed families are shown. Error bars indicate standard deviations.

vulgare genome contains more annotated genes than that of *A. nasatum*, this difference does not contribute much to the observed difference in the genome size of *A. vulgare* relative to *A. nasatum*. Instead, the large difference in genome size (~500 megabases) between the two closely related terrestrial isopods *A. vulgare* and *A. nasatum* can essentially be explained by a difference in TE content likely caused by a higher transposition activity in *A. vulgare* than in *A. nasatum*, subsequent to the divergence between the two species ~25 million years ago. Differential transpositional activity of the same TE families apparently underlies the variation in TE copy numbers and genome size between the two species. Our results highlight the important impact of TEs on the structure and evolution of genomes previously observed in a broad range of organisms, and extended here to two closely related crustacean species.

CRediT authorship contribution statement

Thomas Becking: Conceptualization, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Clément Gilbert:** Conceptualization, Methodology, Supervision, Writing - original draft, Writing - review & editing. **Richard Cordaux:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

This work was funded by Agence Nationale de la Recherche Grant ANR-15-CE32-0006-01 (CytoSexDet) to R.C.

References

[1] T.R. Gregory, J.A. Nicol, H. Tamm, B. Kullman, K. Kullman, I.J. Leitch, B.G. Murray, D.F. Kapraun, J. Greilhuber, M.D. Bennett, Eukaryotic genome size databases, Nucleic Acids Res. 35 (2007) D332–D338, https://doi.org/10.1093/nar/gkl828.
[2] D.A. Petrov, Evolution of genome size: new approaches to an old problem, Trends Genet. 17 (2001) 23–28, https://doi.org/10.1016/S0168-9525(00)02157-0.
[3] T.R. Gregory, The Evolution of the Genome, Academic Press, Burlington, MA, 2005.
[4] M. Lynch, The Origins of Genome Architecture, first ed., OUP USA, Sunderland, Mass, 2007.
[5] S. Linquist, K. Cottenie, T.A. Elliott, B. Saylor, S.C. Kremer, T.R. Gregory, Applying ecological models to communities of genetic elements: the case of neutral theory, Mol. Ecol. 24 (2015) 3232–3242, https://doi.org/10.1111/mec.13219.
[6] A. Canapa, M. Barucca, M.A. Biscotti, M. Forconi, E. Olmo, Transposons, genome size, and evolutionary insights in animals, cytogenet, Genome Res. 147 (2015) 217–239, https://doi.org/10.1159/000444429.
[7] T.A. Elliott, T.R. Gregory, What's in a genome? The C-value enigma and the evolution of eukaryotic genome content, Philos. Trans. R. Soc. B Biol. Sci. 370 (2015) 20140331, https://doi.org/10.1098/rstb.2014.0331.
[8] T.A. Elliott, T.R. Gregory, Do larger genomes contain more diverse transposable elements? BMC Evol. Biol. 15 (2015) 69, https://doi.org/10.1186/s12862-015-0339-8.
[9] A. Kapusta, A. Suh, C. Feschotte, Dynamics of genome size evolution in birds and mammals, Proc. Natl. Acad. Sci. Unit. States Am. 114 (2017) E1460–E1469, https://doi.org/10.1073/pnas.1616702114.
[10] D.A. Petrov, E.R. Lozovskaya, D.L. Hartl, High intrinsic rate of DNA loss in Drosophila, Nature 384 (1996) 346–349, https://doi.org/10.1038/384346a0.
[11] X. Wang, X. Fang, P. Yang, X. Jiang, F. Jiang, D. Zhao, B. Li, F. Cui, J. Wei, C. Ma, Y. Wang, J. He, Y. Luo, Z. Wang, X. Guo, W. Guo, X. Wang, Y. Zhang, M. Yang, S. Hao, B. Chen, Z. Ma, D. Yu, Z. Xiong, Y. Zhu, D. Fan, L. Han, B. Wang, Y. Chen, J. Wang, L. Yang, W. Zhao, Y. Feng, G. Chen, J. Lian, Q. Li, Z. Huang, X. Yao, N. Lv, G. Zhang, Y. Li, J. Wang, J. Wang, B. Zhu, L. Kang, The locust genome provides insight into swarm formation and long-distance flight, Nat. Commun. 5 (2014) 1–9, https://doi.org/10.1038/ncomms3957.
[12] C. Sessegolo, N. Burlet, A. Haudry, Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies, Biol. Lett. 12 (2016) 20160407, https://doi.org/10.1098/rsbl.2016.0407.
[13] V. Talla, A. Suh, F. Kalsoom, V. Dincă, R. Vila, M. Friberg, C. Wiklund, N. Backström, Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (leptidea) butterflies, Genome Biol. Evol. 9 (2017) 2491–2505, https://doi.org/10.1093/gbe/evx163.
[14] S.S. Lower, J.S. Johnston, K.F. Stanger-Hall, C.E. Hjelmen, S.J. Hanrahan, K. Korunes, D. Hall, Genome size in North American fireflies: substantial variation likely driven by neutral processes, Genome Biol. Evol. 9 (2017) 1499–1512, https://doi.org/10.1093/gbe/evx097.
[15] M. Petersen, D. Armisén, R.A. Gibbs, L. Hering, A. Khila, G. Mayer, S. Richards, O. Niehuis, B. Misof, Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects, BMC Evol. Biol. 19 (2019) 11, https://doi.org/10.1186/s12862-018-1324-9.
[16] M. Wu, J.L. Kostyun, L.C. Moyle, Genome sequence of Jaltomata addresses rapid reproductive trait evolution and enhances comparative genomics in the hyper-diverse solanaceae, Genome Biol. Evol. 11 (2019) 335–349, https://doi.org/10.1093/gbe/evy274.
[17] C. Vitte, O. Panaud, H. Quesneville, LTR retrotransposons in rice (Oryza sativa, L.): recent burst amplifications followed by rapid DNA loss, BMC Genom. 8 (2007) 218, https://doi.org/10.1186/1471-2164-8-218.
[18] J.S. Hawkins, S.R. Proulx, R.A. Rapp, J.F. Wendel, Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants, Proc. Natl. Acad. Sci. Unit. States Am. 106 (2009) 17811–17816, https://doi.org/10.1073/pnas.0904339106.
[19] Z.-Q. Zhang, Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness (Addenda 2013), Zootaxa 3703 (2013) 1–82.
[20] G. Giribet, G.D. Edgecombe, Reevaluating the arthropod tree of life, Annu. Rev. Entomol. 57 (2012) 167–186, https://doi.org/10.1146/annurev-ento-120710-100659.
[21] M.A. Chebbi, T. Becking, B. Moumen, I. Giraud, C. Gilbert, J. Peccoud, R. Cordaux, The genome of Armadillidium vulgare (Crustacea, isopoda) provides insights into sex chromosome evolution in the context of cytoplasmic sex determination, Mol. Biol. Evol. 36 (2019) 727–741, https://doi.org/10.1093/molbev/msz010.
[22] T. Becking, M.A. Chebbi, I. Giraud, B. Moumen, T. Laverré, Y. Caubet, J. Peccoud, C. Gilbert, R. Cordaux, Sex chromosomes control vertical transmission of feminizing Wolbachia symbionts in an isopod, PLoS Biol. 17 (2019) e3000438, , https://doi.org/10.1371/journal.pbio.3000438.
[23] T. Becking, I. Giraud, M. Raimond, B. Moumen, C. Chandler, R. Cordaux, C. Gilbert, Diversity and evolution of sex determination systems in terrestrial isopods, Sci. Rep. 7 (2017) 1084, https://doi.org/10.1038/s41598-017-01195-4.
[24] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (2015) 3210–3212, https://doi.org/10.1093/bioinformatics/btv351.
[25] B. Valdebenito-Maturana, G. Riadi, TEcandidates: prediction of genomic origin of expressed transposable elements using RNA-seq data, Bioinformatics 34 (2018) 3915–3916, https://doi.org/10.1093/bioinformatics/bty423.
[26] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, M.D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R.D. Leduc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, Nat. Protoc. 8 (2013) 1494–1512, https://doi.org/10.1038/nprot.2013.084.
[27] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (2010) 841–842, https://doi.org/10.1093/bioinformatics/btq033.
[28] J. Romiguier, P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, E. Loire, J.M. Lourenco, B. Nabholz, C. Roux, G. Tsagkogeorga, A.a.-T. Weber, L.A. Weinert, K. Belkhir, N. Bierne, S. Glémin, N. Galtier, Comparative population genomics in animals uncovers the determinants of genetic diversity, Nature 515 (2014) 261–263, https://doi.org/10.1038/nature13685.
[29] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359, https://doi.org/10.1038/nmeth.1923.
[30] T. Carver, S.R. Harris, M. Berriman, J. Parkhill, J.A. McQuillan, Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data, Bioinformatics 28 (2012) 464–469, https://doi.org/10.1093/bioinformatics/btr703.
[31] B. Czech, G.J. Hannon, One loop to rule them all: the ping-pong cycle and piRNA-guided silencing, Trends Biochem. Sci. 41 (2016) 324–337, https://doi.org/10.1016/j.tibs.2015.12.008.
[32] S.H. Lewis, K.A. Quarles, Y. Yang, M. Tanguy, L. Frézal, S.A. Smith, P.P. Sharma, R. Cordaux, C. Gilbert, I. Giraud, D.H. Collins, P.D. Zamore, E.A. Miska, P. Sarkies, F.M. Jiggins, Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements, Nat. Ecol. Evol. 2 (2018) 174–181, https://doi.org/10.1038/s41559-017-0403-4.
[33] Ö. Deniz, J.M. Frost, M.R. Branco, Regulation of transposable elements by DNA modifications, Nat. Rev. Genet. 20 (2019) 417–431, https://doi.org/10.1038/s41576-019-0106-6.
[34] S. Aparicio, J. Chapman, E. Stupka, N. Putnam, J. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M.D.S. Gelpke, J. Roach, T. Oh, I.Y. Ho, M. Wong, C. Detter, F. Verhoef, P. Predki, A. Tay, S. Lucas, P. Richardson, S.F. Smith, M.S. Clark, Y.J.K. Edwards, N. Doggett, A. Zharkikh, S.V. Tavtigian, D. Pruss, M. Barnstead, C. Evans, H. Baden, J. Powell, G. Glusman, L. Rowen, L. Hood, Y.H. Tan, G. Elgar, T. Hawkins, B. Venkatesh, D. Rokhsar, S. Brenner, Whole-genome shotgun assembly and analysis of the genome of fugu rubripes, Science 297 (2002) 1301–1310, https://doi.org/10.1126/science.1072104.
[35] E. Blass, M. Bell, S. Boissinot, Accumulation and rapid decay of non-LTR retrotransposons in the genome of the three-spine stickleback, Genome Biol. Evol. 4 (2012) 687–702, https://doi.org/10.1093/gbe/evs044.
[36] M. Naville, S. Henriet, I. Warren, S. Sumic, M. Reeve, J.-N. Volff, D. Chourrout, Massive changes of genome size driven by expansions of non-autonomous transposable elements, Curr. Biol. 29 (2019) 1161–1168, https://doi.org/10.1016/j.cub.

2019.01.080 e6.
[37] C.M. Vicient, A. Suoniemi, K. Anamthawat-Jónsson, J. Tanskanen, A. Beharav, E. Nevo, A.H. Schulman, Retrotransposon BARE-1 and its role in genome evolution in the genus hordeum, Plant Cell 11 (1999) 1769–1784, https://doi.org/10.1105/tpc.11.9.1769.
[38] J.L. Bennetzen, J. Ma, K.M. Devos, Mechanisms of recent genome size variation in flowering plants, Ann. Bot. 95 (2005) 127–132, https://doi.org/10.1093/aob/mci008.
[39] B. Nystedt, N.R. Street, A. Wetterbom, A. Zuccolo, Y.-C. Lin, D.G. Scofield, F. Vezzi, N. Delhomme, S. Giacomello, A. Alexeyenko, R. Vicedomini, K. Sahlin, E. Sherwood, M. Elfstrand, L. Gramzow, K. Holmberg, J. Hällman, O. Keech, L. Klasson, M. Koriabine, M. Kucukoglu, M. Käller, J. Luthman, F. Lysholm, T. Niittylä, Å. Olson, N. Rilakovic, C. Ritland, J.A. Rosselló, J. Sena, T. Svensson, C. Talavera-López, G. Theißen, H. Tuominen, K. Vanneste, Z.-Q. Wu, B. Zhang, P. Zerbe, L. Arvestad, R. Bhalerao, J. Bohlmann, J. Bousquet, R.G. Gil, T.R. Hvidsten, P. de Jong, J. MacKay, M. Morgante, K. Ritland, B. Sundberg, S.L. Thompson, Y.V. de Peer, B. Andersson, O. Nilsson, P.K. Ingvarsson, J. Lundeberg, S. Jansson, The Norway spruce genome sequence and conifer genome evolution, Nature 497 (2013) 579–584, https://doi.org/10.1038/nature12211.
[40] M. Dupeyron, S. Leclercq, N. Cerveau, D. Bouchon, C. Gilbert, Horizontal transfer of transposons between and within crustaceans and insects, Mobile DNA 5 (2014) 4, https://doi.org/10.1186/1759-8753-5-4.