Evolutionary reshuffling in the Errantivirus lineage Elbe within the *Beta vulgaris* genome

Cora Wollrab¹, Tony Heitkam¹², Daniela Holtgräwe², Bernd Weisshaar³, André E. Minoche⁴, Juliane C. Dohm⁴, Heinz Himmelbauer⁴ and Thomas Schmidt¹* ¹Department of Biology, Dresden University of Technology, D-01062, Dresden, Germany, ²Institut de Biologie de l’Ecole Normale Supérieure (IBENS), Paris, France, ³Department of Biology, University of Bielefeld, D-33594 Bielefeld, Germany, and ⁴Centre for Genomic Regulation (CRG)-Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain

Received 12 April 2012; revised 2 July 2012; accepted 11 July 2012; published online 24 September 2012.

*For correspondence (e-mail thomas.schmidt@tu-dresden.de).

SUMMARY

LTR retrotransposons and retroviruses are closely related. Although a viral *envelope* gene is found in some LTR retrotransposons and all retroviruses, only the latter show infectivity. The identification of Ty3-gypsy-like retrotransposons possessing putative *envelope*-like open reading frames blurred the taxonomical borders and led to the establishment of the Errantivirus, Metavirus and Chromovirus genera within the *Metaviridae*. Only a few plant Errantiviruses have been described, and their evolutionary history is not well understood. In this study, we investigated 27 retroelements of four abundant Elbe retrotransposon families belonging to the Errantiviruses in *Beta vulgaris* (sugar beet). Retroelements of the Elbe lineage integrated between 0.02 and 5.59 million years ago, and show family-specific variations in autonomy and degree of rearrangements: while Elbe3 members are highly fragmented, often truncated and present in a high number of solo LTRs, Elbe2 members are mainly autonomous. We observed extensive reshuffling of structural motifs across families, leading to the formation of new retrotransposon families. Elbe retrotransposons harbor a typical *envelope*-like gene, often encoding transmembrane domains. During the course of Elbe evolution, the additional open reading frames have been strongly modified or independently acquired. Taken together, the Elbe lineage serves as retrotransposon model reflecting the various stages in Errantivirus evolution, and allows a detailed analysis of retrotransposon family formation.

Keywords: Errantivirus, LTR retrotransposon, Ty3-gypsy, *Beta vulgaris*, envelope-like open reading frame.

INTRODUCTION

Present in high copy numbers and ubiquitous across all eukaryotes, retrotransposons play a major role in the maintenance of genome structure, function and evolution (Kumar and Bennetzen, 1999). Based on the presence of flanking long terminal repeats (LTRs), they are divided into LTR and non-LTR retrotransposons. Members of both subclasses replicate by reverse transcription of their mRNA and subsequent integration of the resulting cDNA into other genomic positions (Boeke and Corces, 1989). Thus, both their copy number and the host genome size increase (Bennetzen, 2002; Ma and Bennetzen, 2004). Plant genomes in particular tend to accumulate retroelements in high copy numbers; in maize, for example, LTR retrotransposons account for nearly 75% of the genome (Baum et al., 2009).

To prevent severe damage of the host genome, mechanisms for suppression of LTR retrotransposon proliferation have evolved. First, heterochromatic marks silence retrotransposons transcription epigenetically (Slotkin and Martienssen, 2007). Second, recombination events often lead to deletion of internal sequences and thus formation of non-autonomous LTR retrotransposons or solo LTRs, and hence counteract the increase in genome size (Devos et al., 2002; Ma et al., 2004).

The essential gene products for proliferation of LTR retrotransposons are encoded by the open reading frames (ORFs) designated *gag* and *pol*. Based on the order of the enzymatic domains of the aspartic protease (AP), reverse transcriptase (RT), RNase H (RH) and integrase (IN), LTR retrotransposons are classified either as Ty3-gypsy-like (*Metaviridae*) or Ty1-copia-like (*Pseudoviridae*) type (Kumar and Bennetzen, 1999; Hull, 2001). Retroviruses and LTR retrotransposons are evolutionarily closely related (Eickbush, 1994; Eickbush and Malik, 2002). In
particular, Ty3-gypsy proteins show high similarity to their retroviral equivalents (Xiong and Eickbush, 1990); however, they lack the infectivity-facilitating Envelope (Env) proteins. The identification of LTR retroelements with an additional env ORF blurred the taxonomical borders between retroviruses and LTR retrotransposons (Song et al., 1994; Wright and Voytas, 1998). Therefore, the Errantivirus, Metavirus and Chromovirus genera within the family of Ty3-gypsy-like retrotransposons were established (Fauquet and Mayo, 2001). In plants, Athila from Arabidopsis thaliana was the first Errantivirus discovered (Pelissier et al., 1995; Wright and Voytas, 1998). However, infectivity of Athila has not yet been demonstrated. Cross-species analyses of Athila RT motifs revealed high heterogeneity of the family members and the presence of Athila relatives in both monocots and dicots (Wright and Voytas, 1998, 2002; Marco and Marín, 2005). SIRE1, a representative of the Ty1-copia-like retrotransposons containing an env ORF has been identified in soybean (Laten et al., 1998), and was namesake of the genus Sireviruses (Fauquet, 2005).

Env glycoproteins enable a virus particle to infect another cell by mediating receptor binding, adsorption by the host cell, and particle budding (Gallo et al., 2003). However, Env proteins are extremely variable and only weakly conserved, complicating identification of the corresponding ORFs from sequence data. Usually, Env proteins comprise two subunits: the larger surface protein that enables viral adsorption by binding specific cell surface receptors, and the transmembrane (TM) protein that mediates the entry into the host cell by cell membrane fusion and allows transfer between cells (Coffin, 1990).

Sugar beet (Beta vulgaris) possesses a relatively small genome of 758 Mb (Arumuganathan and Earle, 1991) and contains at least 63% repetitive DNA (Flavell et al., 1974; Menzel et al., 2008). A major proportion of these repeats has been characterized in cultivated and wild beets of the genera Beta and Patellifolia (Schmidt and Metzlaff, 1991; Schmidt et al., 1991; Dechyeva and Schmidt, 2006; Zakrzewski et al., 2010), and has been found to include a variety of DNA transposons (Jacobs et al., 2004; Menzel et al., 2006) and retrotransposons (Heitkam and Schmidt, 2009; Weber and Schmidt, 2009; Wenke et al., 2009). Recently, Weber et al. (2010) identified a sireviral Ty1-copia retrotransposon Cotzilla with a putative env ORF that lacks regions encoding typical transmembrane domains.

Here, we have analyzed four families of an errantiviral Ty3-gypsy lineage named Elbe in B. vulgaris at various stages of retrotransposon rearrangement that reflect their evolution. We show that Elbe Errantiviruses have a family-specific large-scale structure. Genome-wide analysis provides evidence of extensive recombination and reshuffling between domains of various Elbe retrotransposons, leading to formation of new families and extensive modification or capture of putative env ORFs.

RESULTS

Elbe retrotransposon families constitute a lineage of Errantiviruses

To investigate B. vulgaris Errantiviruses, a 385 bp fragment of the B. vulgaris-RT homologous to the Athila RT was used to detect overlapping B. vulgaris bacterial artificial chromosome (BAC) ends (McGrath et al., 2004). Their subsequent assembly resulted in the creation of a complete in silico retrotransposon. Domain-specific probes were derived from the in silico element (Table S1) and hybridized to a high-density filter containing 9216 BACs to identify full-length copies. Partial sequencing of BAC 69P23 revealed two retrotransposons with characteristic errantiviral gag-pol polyprotein domains. These elements were designated Elbe1 and Elbe2. Using the Elbe2 gag-pol polyprotein as the query for a database search resulted in identification of Elbe3 and Elbe4 within the sequence of BAC 137 (DO374067) from B. vulgaris chromosome 9 (Schulte et al., 2006). The four Elbe retrotransposons differ strongly in size and integrity. Based on their LTR sequence similarity, which is less than 70%, they represent four different but related families of LTR retrotransposons (Wicker et al., 2007). The recent availability of the draft assembly of the B. vulgaris genome sequence allowed expansion of the number of Elbe members. Based on similarity of more than 80% of the LTRs, 23 additional members belonging to the four Elbe families have been identified. Their structural features are shown in Figure 1 and summarized in Table 1.

As typical for LTR retrotransposons, Elbe Errantiviruses are usually flanked by target site duplications (TSDs) of five base pairs. The LTRs of approximately 1400 bp are terminated by the characteristic dinucleotides 5'-TG and CA-3', which are important for integration of retroviral sequences (Temin, 1980). Numerous putative promoter motifs have been identified in the 5' LTRs of Elbe retrotransposons (Figure S1).

All families of the Elbe lineage contain non-autonomous rearranged members that lack an intact gag-pol ORF and/or env-like ORF. Two autonomous Elbe1 and most Elbe2 members are full-length and possess two intact ORFs, but the remaining families mainly comprise structurally defective retroelements, harboring large insertions or deletions within their protein-coding ORFs. Intact Errantiviruses contain a 3' UTR, a 5' UTR, and an internal UTR separating the two ORFs.

Most of the Elbe retrotransposons contain an 18 bp primer binding site that is complementary to the 3' end of the aspartic acid tRNA from Arabidopsis thaliana. Adjacent to the 3' LTR and separated by a thymine dinucleotide spacer as part of the attachment site, a polypurine tract (PPT) was also detected in all Elbe members. Multiple PPTs were identified within the 3' UTR of all Elbe retrotransposons (Figure 5b).

Eight of the 27 Elbe Errantiviruses harbor insertions within various domains. Although three insertions with high
Figure 1. Schematic representation of Errantiviruses of the Elbe lineage. Reference family members Elbe1, Elbe2, Elbe3, and Elbe4 are marked with an asterisk. Open terminal arrows represent the flanking LTRs, with vertical wavy lines indicating incomplete LTRs due to missing sequence data. Rectangles indicate the gag-pol ORF (gray) and the family-specific putative env ORF (hatched). Conserved domains are shown: gag, protease (AP), reverse transcriptase (RT), RNase H (RH), integrase (IN). Vertical bars represent the primer binding site (black, adjacent to the 5' LTR) and single or multiple PPTs (gray, adjacent to the 3' LTR). Insertions are indicated as white triangles, with size, sequence type and relative orientation shown in dashed boxes.
<table>
<thead>
<tr>
<th>Element</th>
<th>Length (bp)</th>
<th>TSD</th>
<th>LTR length (5'-/3', bp)</th>
<th>Primer binding site</th>
<th>UTR length (5'-/internal/3', bp)</th>
<th>gag–pol ORF length (bp)</th>
<th>ORF env-like length (bp)</th>
<th>Number of PPTs</th>
<th>Integration-time (MYA)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elbe1</td>
<td>8019</td>
<td>CTGAC</td>
<td>1439/1437</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>138/-/2473</td>
<td>2532*</td>
<td>843*</td>
<td>–</td>
<td>4</td>
<td>0.09</td>
</tr>
<tr>
<td>Elbe1-1</td>
<td>7199</td>
<td>TCACC</td>
<td>1501/1505</td>
<td>TGGCACCTGTTGCCGGGGA</td>
<td>144/-/1660</td>
<td>2389*</td>
<td>792*</td>
<td>–</td>
<td>3</td>
<td>2.77</td>
</tr>
<tr>
<td>Elbe1-2</td>
<td>11 820</td>
<td>–</td>
<td>965/-1449</td>
<td>TTGTCGCCGTTGCCGGGGA</td>
<td>141/860/1602</td>
<td>5412</td>
<td>1798</td>
<td>1391/462</td>
<td>2</td>
<td>1.34d</td>
</tr>
<tr>
<td>Elbe1-3</td>
<td>4770b</td>
<td>AATA(A/T)</td>
<td>1548/1489</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>154/-/1699b</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3</td>
<td>5.59</td>
</tr>
<tr>
<td>Elbe1-4</td>
<td>11 889</td>
<td>TCCAT</td>
<td>1009/1503</td>
<td>–</td>
<td>100/-/1919/1603</td>
<td>5377</td>
<td>1784</td>
<td>1377/457</td>
<td>2</td>
<td>2.66d</td>
</tr>
<tr>
<td>Elbe2</td>
<td>12 456b</td>
<td>CAAG</td>
<td>1470/1472</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>303/949/1558</td>
<td>5297*</td>
<td>1764*</td>
<td>1437/478</td>
<td>3</td>
<td>0.14</td>
</tr>
<tr>
<td>Elbe2-1</td>
<td>12 527</td>
<td>AGCAC</td>
<td>1473/1471</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>295/997/1431</td>
<td>5301</td>
<td>1765</td>
<td>1431/476</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>Elbe2-2</td>
<td>12 553</td>
<td>CTATC</td>
<td>1463/1443</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>315/1020/1578</td>
<td>5283</td>
<td>1750</td>
<td>1451/479</td>
<td>3</td>
<td>2.42</td>
</tr>
<tr>
<td>Elbe2-3</td>
<td>11 424</td>
<td>CCTT(T/C)</td>
<td>1448/1447</td>
<td>TGGACCTGTTGCCGGGGA</td>
<td>309/1001/1575</td>
<td>4690*</td>
<td>1557a</td>
<td>954*/317*</td>
<td>3</td>
<td>2.51</td>
</tr>
<tr>
<td>Elbe2-4</td>
<td>11 705</td>
<td>–</td>
<td>1483/1158a</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>295/899/1466</td>
<td>5227</td>
<td>1735</td>
<td>1122/391</td>
<td>3</td>
<td>4.19d</td>
</tr>
<tr>
<td>Elbe2-5</td>
<td>12 244</td>
<td>GAA(T/A)C</td>
<td>1312*/1527</td>
<td>TGGCGCCATGCAGGGGA</td>
<td>309/953/1474</td>
<td>5274</td>
<td>1747</td>
<td>1512/461</td>
<td>2*</td>
<td>4.03d</td>
</tr>
<tr>
<td>Elbe2-6</td>
<td>12 326b</td>
<td>CATGA</td>
<td>1526/1501</td>
<td>TGGCGCCATGCAGGGGA</td>
<td>310/854/1503</td>
<td>5192b</td>
<td>1714b</td>
<td>1128/475b</td>
<td>3</td>
<td>3.61</td>
</tr>
<tr>
<td>Elbe3</td>
<td>6467</td>
<td>A(A/T)ACC</td>
<td>1358/1365</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>129/-/2631</td>
<td>–</td>
<td>894</td>
<td>327</td>
<td>4</td>
<td>1.00</td>
</tr>
<tr>
<td>Elbe3-1</td>
<td>4622</td>
<td>–</td>
<td>1185/-1348</td>
<td>–</td>
<td>–/-/2089</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>4</td>
<td>0.74d</td>
</tr>
<tr>
<td>Elbe3-2</td>
<td>7185b</td>
<td>GTACA</td>
<td>1369/-1367</td>
<td>TGGACGCATTGCCGGGGA</td>
<td>115/-/3316*</td>
<td>–</td>
<td>1018</td>
<td>336</td>
<td>4</td>
<td>2.64</td>
</tr>
<tr>
<td>Elbe3-3</td>
<td>6492</td>
<td>GTCGT</td>
<td>1356/-1225</td>
<td>TGGCGCTGTTGCCGGGGA</td>
<td>274/458/2796</td>
<td>219*</td>
<td>71*</td>
<td>184*/55*</td>
<td>4</td>
<td>2.27d</td>
</tr>
<tr>
<td>Elbe3-4</td>
<td>3919</td>
<td>–</td>
<td>1430/-382*</td>
<td>TGGCGCGTGCAGGGGA</td>
<td>63/-/1444</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3*</td>
<td>1.41d</td>
</tr>
<tr>
<td>Elbe3-5</td>
<td>5184</td>
<td>–</td>
<td>658/-1356</td>
<td>TGGCGCTGTTGCCGGGGA</td>
<td>381/234/2096</td>
<td>150*</td>
<td>50*</td>
<td>319*/107*</td>
<td>4</td>
<td>3.07d</td>
</tr>
<tr>
<td>Elbe3-6</td>
<td>6691</td>
<td>–</td>
<td>1349/1140</td>
<td>TGGCGCTGTTGCCGGGGA</td>
<td>295/336/2022</td>
<td>180*</td>
<td>60*</td>
<td>369*/123*</td>
<td>4</td>
<td>1.54d</td>
</tr>
<tr>
<td>Elbe4</td>
<td>9999b</td>
<td>GTTGG</td>
<td>1404/1371</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>281/854/1558b</td>
<td>3390</td>
<td>1124*</td>
<td>1139/375</td>
<td>2*</td>
<td>2.28</td>
</tr>
<tr>
<td>Elbe4-1</td>
<td>7810</td>
<td>(A/T)CCAT</td>
<td>1471/1315</td>
<td>TGGTGCCGTTGCCGGGGA</td>
<td>92/1249/1498</td>
<td>820*</td>
<td>417*</td>
<td>1365/550</td>
<td>3</td>
<td>2.06d</td>
</tr>
<tr>
<td>Elbe4-2</td>
<td>7453</td>
<td>ATAC(A/C)</td>
<td>1478/1497</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>280/-/1918</td>
<td>2280*</td>
<td>756*</td>
<td>–</td>
<td>3</td>
<td>2.16</td>
</tr>
<tr>
<td>Elbe4-3</td>
<td>6570</td>
<td>GTTC(T/C)</td>
<td>1524/1526</td>
<td>TGGCGCCGTTGCCGGGGA</td>
<td>281/-/1076</td>
<td>2163*</td>
<td>718*</td>
<td>–</td>
<td>2*</td>
<td>4.32</td>
</tr>
<tr>
<td>Elbe4-4</td>
<td>6941</td>
<td>–</td>
<td>1297/1286</td>
<td>TGGCGCTGTTGCCGGGGA</td>
<td>282/-/1396</td>
<td>2680*</td>
<td>881*</td>
<td>–</td>
<td>2*</td>
<td>2.94</td>
</tr>
<tr>
<td>Elbe4-5</td>
<td>11 373b</td>
<td>A(C)/GICT</td>
<td>1282/1301</td>
<td>TGGACCGTTGCCGGGGA</td>
<td>208/771/1545</td>
<td>4873*</td>
<td>1615b</td>
<td>1321/437</td>
<td>3</td>
<td>4.62</td>
</tr>
<tr>
<td>Elbe4-6</td>
<td>8915b</td>
<td>ACGAT</td>
<td>910/1333</td>
<td>TGGACCGTTGCCGGGGA</td>
<td>751/771/1780</td>
<td>2204*</td>
<td>727*</td>
<td>1166b/382*</td>
<td>3</td>
<td>3.81d</td>
</tr>
</tbody>
</table>

bp, base pairs; AA, amino acids.

*Deletion.

*Insertion has been removed.

*End of contig.

*Age estimation with partial LTR sequences.
similarity to Elbe retrotransposons or parts thereof show
nested organization, five integrations originate from non-
LTR retrotransposons or are of unknown origin (Figure 1). 
Thus, Elbe1 is integrated in reverse orientation into the
reverse transcriptase region of Elbe2, while Elbe3 is inte-
grated in reverse orientation into the PPT of Elbe4. Three
additional integrations are flanked by the inverted dinucleo-
tides (TG/CA) and a 5 bp TSD, and thus are considered as
solo LTRs from as yet unknown LTR retrotransposons.

Comparative analyses of Elbe RT amino acid sequences
with RTs of other eukaryotic retroelements (Xiong and 
Eickbush, 1990) grouped Elbe elements within the Erranti-
viridae clade defined by Athila4-2 (Wright and Voytas, 2002)
(Figure 2). Conserved amino acid motifs of the gag-pol
polyprotein, such as the zinc finger domain of the gag
region, the seven conserved RT domains (Xiong and Eick-
bush, 1990) and the three integrase domains, including the
zinc finger, the DD35E domain (Fayet et al., 1999) and the
GPY/F motif (Malik and Eickbush, 1999), confirm the assign-
ment to the errantiviral clade (Figure S2).

env-like ORFs in the Elbe Errantiviruses

A putative env ORF downstream of the gag-pol polyprotein
is characteristic of Elbe retrotransposons. The predicted
proteins are approximately 400 amino acids long
(Table 1), and show remarkable differences in amino acid
composition, supporting the assignment of Elbe retro-
transposons to four families (Figure S3). The ability of
retroviruses to penetrate cell membranes is often associated
with a change in the hydrophobicity of retroviral Env pro-
teins. Hydrophobic signatures typical of TM domains were
predicted in the env-like ORF by TMpred and TMHMM
(Hofmann and Stoffel, 1993; Krogh et al., 2001) (Figure S4).
Both programs detected TM domains with significant
scores, but with remarkable differences in position and
number. In particular, the TM domain of Elbe3 is charac-
terized by high scores and probability values (Figure 3a).
The Env protein of the Elbe3 family is significantly different
from the Env protein of Elbe1, Elbe2 and Elbe4, which also contain
regions of high homology but form independent branches in
the neighbor-joining tree (Figure 3b). This indicates either
severe mutation and rearrangements, or, more likely, independent acquisition of env-like ORFs during evolution.

Reshuffling of retrotransposon families in the Elbe
Errantivirus lineage

Commonly, retrotransposons are grouped according to
conservation of LTR sequences as shown for the 5’ LTR in
Figure 4a. However, due to family-specific signatures, 3’
UTRs may also be used for classification. Surprisingly,
comparison of Elbe 3’ UTRs resulted in an inconsistent
grouping of Elbe1 members within the Elbe2 and Elbe3
families (bold in Figure 4b), which is also supported by low
conservation in pairwise Elbe1 3’ UTR sequence compari-
sions (Figure 4c). The 3’ UTRs of Elbe1-1 and Elbe1-3 show
high homology (75–79%) to the 3’ UTR of Elbe2 copies, al-
though the homology of the remaining regions confirms the
clear assignment to the Elbe1 family. Similarly, Elbe1 shows
between 75 and 79% identity to the 3’ UTR of the Elbe3
elements, thus suggesting recombination of the 3’ UTRs
between retrotransposons of different families.

Figure 5(a) schematically represents the three recombi-
ned Elbe1 retrotransposons Elbe1, Elbe1-1, and Elbe1-3, including
the identity values of the corresponding regions. Frequent
recombination is strongly confirmed by the number of PPTs
within the 3’ UTR of all Elbe members (Figure 5b). Although
Elbe2 and Elbe4 retrotransposons have three PPTs, and the
Elbe3 family have four PPTs, members of the Elbe1 family vary
and contain up to four PPTs in the reshuffled 3’ UTR regions.
The distances between the most upstream and the most
downstream PPT are similar in all 3’ UTRs, indicating a
common origin. Most PPTs have a conserved upstream
dinucleotide motif (TT), while the two nucleotides down-
stream are diverged. However, only the PPT directly upstream
of the 3’ LTR contains the nucleotides TT adjacent to the 3’ LTR,
suggesting a functional role within plus-stand synthesis by
defining an attachment site essential for integrase binding.

Structural analysis of all LTR-flanked Elbe members
revealed a high degree of rearrangements within and
between families (Figures 1 and 5). In order to quantify this

Figure 2. Classification of Elbe members. The relationship of Elbe retrotransposons to other Errantiviruses is visualized in a dendrogram based on an RT amino acid sequence alignment. The following retrotransposons are included: Rire1 (Oryza australiensis); Cire1 (Citrus sinensis); SALIRE1, Cotzilla1 and Cotzilla3 (Beta vulgaris); Tnt1 and Tnt1 (Nicotiana tabacum); SIRE1, Diaspora and Calypso (Glycine max); Hopie, Opie2 and PREM-2 (Zea mays); Athila4-2 (Arabidopsis thaliana); Bagy2-1 (Hordeum vulgare); Cyclops-2 (Pisum sativum); Beetlet1 and Beetlet2 (Patellifolia comumens). The retro-
transposon gypsy (Drosophila melanogaster) and the retrovirus Osvaldo (Drosophila buzzanti) are also included. Bootstrap values are indicated as a
percentage of 1000 replicates.
reshuffling on a genome-wide scale, we hybridized ten Elbe sequences to 9216 BACs on a high-density filter containing 1.5 B. vulgaris genome equivalents. LTR regions of the four Elbe families as well as all internal Elbe2 protein domains (gag, AP, RT, RH, IN, env) served as probes. We identified 402 BACs containing at least one complete or partial Elbe element (Figure 6a, right panel). In total, 93 hybridization patterns were detected, demonstrating the high degree of rearrangements and internal deletions within Elbe members. A total of 116 BACs showed hybridization signals with all seven Elbe2-specific probes, whereas 52 BACs contained only truncated Elbe2 copies. In contrast, 121 BACs contain only parts of the gag–pol polyprotein of the Elbe2 family and no Elbe2 LTR, while 344 of the 402 BACs possess at least one Elbe LTR. Noteworthy is the detection of solo LTRs, mostly derived from Elbe3, for which the LTR was exclusively detected on 65 BACs.

In order to quantify the occurrence of solo LTRs within the B. vulgaris genome, a local BLAST search was performed using the 5’ LTRs of retrotransposons of the Elbe lineage (Figure 6b). Solo LTRs of all Elbe families with and without TSDs were detected. Due to truncations at the 5’ or 3’ end, many solo LTRs showed only a regional and hence partial sequence homology to the query. The Elbe3 family shows the highest number of solo LTRs (200), confirming the results of the hybridization experiment (Figure 6a). There were similar numbers with and without TSDs; however, TSD-flanked solo LTRs showed a preference for strong

Figure 3. Variable sequence and structure of Elbe Env-like proteins. 
(a) Putative transmembrane (TM) domains in the Env amino acid sequences of representative members of the Elbe lineage. The TMHMM output shows the probability of TM domains (black, left y axis). TMpred generates scores based on the likelihood of encoded TM domains, whereby only scores greater than 500 (dashed line) are considered significant for a TM domain (dark gray, inside → outside; light gray, outside → inside, right y axis). The x axis indicates amino acid sequence position. 
(b) Dendrogram representing alignments of the amino acid sequence of the Env-like proteins. All Elbe elements possessing an env ORF were analyzed. For comparison, the Env-like proteins of the following retrotransposons are included: gypsy (Drosophila melanogaster), Calypso (Glycine max), Athila4-2 (Arabidopsis thaliana), Cyclops-2 (Pisum sativum) and Cotzilla1 (Beta vulgaris). Bootstrap values are indicated as a percentage of 1000 replicates.
sequence conservation over their full length. The high number of solo LTRs with only regional homology strongly suggests recombination between LTRs. The total number of solo LTRs of the remaining Elbe families was considerably lower, indicating a decreased recombination frequency compared with the Elbe3 family.

The similarity of UTRs, the number of PPTs, the distribution in genomic sequences and the high number of conserved or truncated solo LTRs strongly suggest that the various families of the Elbe lineage evolved by recombination and reshuffling of Elbe precursor retrotransposons.

**Figure 4.** Reshuffling of retrotransposons of the Elbe Errantivirus lineage. Neighbor-joining tree based on (a) the sequence of the 5' LTR, and (b) the sequence of the 3' UTR of Elbe Errantiviruses. Bootstrap values are indicated as percentage of 1000 replicates. (c) Pairwise sequence identity matrix of the 3' UTRs of the Elbe retrotransposons show higher similarity across families. The numbers indicate the nucleotide sequence identity (percentages). Gray shading indicates different homology values.

**Proliferation of the Elbe Errantivirus lineage in the genera Beta and Patellifolia**

Estimation of the transposition time revealed large differences among and particularly within individual Elbe families (Table 1). Elbe2 was active 0.02 million years ago (MYA) while Elbe2-4 transposed 4.19 MYA. A similar variation has been calculated for the Elbe1 copies (0.09 MYA for Elbe1 versus 5.59 MYA for Elbe1-3).

The abundance and organization of the Elbe Errantiviruses within the best genera Beta (sections Beta,
Reshuffling of Errantiviruses

Figure 5. Rearrangement of Elbe Errantiviruses.
(a) Recombination of the 3’ UTR of Elbe1-1, Elbe1-3 and Elbe1 across different Elbe families. Arrows and boxes correspond to those shown in Figure 1. Sequence similarities of homologous regions are indicated in the shaded areas.
(b) Comparison of the polypurine tracts (PPTs) within the 3’ UTR of the Elbe retrotransposons. Black shading indicates identity of 100%. The numbers indicate the distance (nucleotides) between the PPTs. The start of the 3’ LTR is shown in gray.

Corollinae and Nanae), Patellifolia and outgroup Spinacia oleracea were analyzed by Southern hybridization using LTR-derived probes (467–627 bp) and env-specific probes (226–482 bp, Table S1) of retrotransposons of the Elbe lineage (Figure 7). Highly abundant copies with many conserved restriction sites exist in all Beta species as detected by strong signals and a conserved banding pattern (Figure 7a–d). In particular, hybridization with the Elbe3 LTR reveals very strong signals, suggesting the occurrence of many retroelement copies and solo LTRs. Although Elbe1 and Elbe4 show strong signals in the related genus Patellifolia, only weak signals were detected with Elbe2 and Elbe3 probes. Elbe3 also was detectable in Spinacia oleracea. Compared to the LTRs, the Elbe env-like ORFs show a similar hybridization pattern with weaker signal strength (Figure 7e–h). Although strongly rearranged, Elbe3 is the most abundant family. Although Elbe3 LTR signals occur in S. oleracea, no env signal was
Figure 6. Genome-wide analyses of recombination hot spots and solo LTR formation.
(a) Existence of Elbe probe hybridizations detected in sugar beet BACs. Of 9216 BACs (1.5-fold genome coverage), 402 showed at least one signal after hybridization with one of ten Elbe probes. The probes were constructed to detect the family-specific LTRs of all four Elbe families as well as internal domains of Elbe2. Ninety-three different hybridization patterns were observed, and positive hybridization signals are shown in dark gray. On the right, the number of detected BACs with the respective hybridization pattern is plotted using light gray bars.
(b) Family-specific numbers of Elbe solo LTRs within the genome of Beta vulgaris. The histogram shows solo LTRs of the four Elbe families with target site duplication (TSD, black bars) and without TSD (gray bars). Solo LTRs that are completely homologous, not homologous at the 5′ end or not homologous at the 3′ end are indicated by rectangles, circles and triangles, respectively.
detected, indicating the absence or strong sequence divergence of the env-like ORF in this distantly related species.

Physical mapping of the Elbe retrotransposons on B. vulgaris chromosomes was performed by fluorescent in situ hybridization (FISH). LTR probes were hybridized to mitotic metaphase, prometaphase and interphase nuclei, and revealed dispersed signals on all chromosomes (Figure 8). Elbe Errantiviruses are organized in clusters in heterochromatic regions, mostly visible as brightly stained DAPI-positive signals. FISH on interphase nuclei confirmed that some heterochromatic centers lack a hybridization signal (Figure 8b,d,f,h). Elbe retrotransposons show a strong bias to subterminal integration, and are largely excluded from many centromeric and pericentromeric regions. The variable signal strength along chromosomes is most likely explained by hybridization to multiple Elbe copies in a nested organization or in close vicinity and below the threshold of resolution. This is in line with the nested pattern for an Elbe1 copy (inserted in Elbe2) and an Elbe3 copy (inserted in Elbe 4), as shown in Figure 1.

**DISCUSSION**

We have identified an Errantivirus lineage containing the retrotransposon families Elbe1, Elbe2, Elbe3 and Elbe4 that is widespread in the genus Beta and is also detected in the genus Patellifolia. Their high abundance facilitates recombinations leading to multiple rearrangements, formation of non-autonomous copies, and formation of related retrotransposon families. We detected various levels of reshuffling: (i) nested integration and truncation of retrotransposons, (ii) exchange of the 3′ UTR between retrotransposons of different families, and (iii) a high number of partially rearranged solo LTRs.

**Structural diversity, organization and rearrangements of Elbe retrotransposon families**

Illegitimate recombination counteracts retrotransposon proliferation and produces non-autonomous copies by deletion of internal sequences (Devos et al., 2002). Three Elbe families contain autonomous and non-autonomous members, and the extent of deletions and rearrangements is different for each family. The Elbe2 retrotransposons predominantly consist of autonomous elements, but the identified Elbe3 members exclusively are non-autonomous (Figure 1).

An essential feature of autonomous and non-autonomous retrotransposon partners is the conservation of the LTR sequences and adjacent UTRs, most likely due to co-evolution of the partners in a host genome (Jiang et al., 2002; Du et al., 2009). For example, the autonomous and non-autonomous retrotransposons SARE and SNRE in Glycine max share similar LTRs (Du et al., 2009). However, retroelement partners may also belong to different monophyletic groups. In Oryza sativa, the putative
retrotransposon partners Rire2 and Dasheng exclusively group into two highly diverged clades based on their LTR sequences, suggesting only few inter-element recombinations (Jiang et al., 2002). We observed two features of evolution of non-autonomous Elbe copies. First, comparison of Elbe LTRs allowed formal classification into four families, clearly indicating a partnership between autonomous and non-autonomous members of each family (Figure 4a). Second, comparison of the 3’ UTR revealed frequent inter-element recombination events between members of different Elbe families (i.e. between Elbe1 and Elbe2, and between Elbe1 and Elbe3, respectively). In these examples, we observed family formation in progress, but were able to deduce the origin of the individual sequences (Figures 4b,c and 5a). Similar reshuffling of two related lineages was observed in the Errantiviruses of A. thaliana and G. max (Marco and Marin, 2008).

The 3’ UTR comprising the region between the env-like ORF and the 3’ LTR is a hotspot of Elbe recombination. In G. max, recombination occurred upstream of the PPT (Du et al., 2009), indicating a preference for inter-element recombination close to the 3’ LTR. The rearranged retroelements Elbe1, Elbe1-1 and Elbe1-3, as well as the G. max elements, are flanked by TSDs (Table 1). This strongly suggests that the reshuffling occurred during reverse transcription of a novel copy, probably by template switching of the reverse transcriptase, showing that the recombinants are fully mobile and reverse transcription was not aborted. An alternative explanation may be recombination of autonomous retroelements after integration into the host genome.

The number and positions of PPTs served as markers for Elbe family-specific 3’ UTRs (Figure 5b). Multiple PPTs have also been reported for Athila from A. thaliana, Calypso from G. max and Cyclops from P. sativum (Wright and Voytas, 2002). Up to four PPT copies have been identified in the Elbe Errantiviruses, of which the terminal PPT corresponds to a distinctive integrase attachment site. This site attracts binding of the integrase, thus defining the 5’ end of the new retrotransposon copy after plus-strand synthesis, and is therefore essential for successful insertion of the new copy into the host genome (Bowman et al., 1996). Multiple PPTs may originate from rearrangements between Elbe copies followed by selection, because only the terminal PPT would enable full-length synthesis of the plus strand while internally initiated strands would lead to truncations.

Nested integrations are a common feature of LTR retrotransposons (SanMiguel et al., 1996), and it is likely that nested transposition fragmented intact Elbe copies and produced the observed non-autonomous elements. We
identified eight Elbe Errantiviruses containing insertions of complete LINEs, solo LTRs or LTR retrotransposons (Figure 1; e.g. Elbe1 within Elbe2 and Elbe3 within Elbe4). Complete nested Elbe retrotransposons were only detected on partially sequenced BACs and not in the draft genome sequence of *B. vulgaris*. This reflects the limited capability of next-generation sequencing assembly software to resolve segmental duplications and homologous repeats with identities exceeding 85% (Alkan et al., 2011).

The clustered organization of Elbe retrotransposons was also indicated by the dispersed patterns of FISH signals on *B. vulgaris* prometaphase chromosomes (Figure 8). Most Elbe retrotransposons were physically mapped in subterminal chromosome positions, and are largely excluded from many centromeric and pericentromeric regions. This is in contrast to the Errantiviruses Athila from *A. thaliana* and SNARE from *G. max*, which are located in recombination-suppressed pericentromeric regions (Kumekawa et al., 2000; Du et al., 2010).

The occurrence of nested integrations is consistent with the comparative hybridization to *B. vulgaris* BACs showing that 31% of positive BACs (126 of 402) harbor two or more copies of different Elbe families (Figure 8a). We assume that, due to their heterochromatic localization, Elbe retrotransposons provide a safe harbor for novel retrotransposon insertions.

In addition to illegitimate recombination, Elbe Errantiviruses are involved in unequal homologous recombination. This unequal intra-element recombination between the two LTRs of a single retroelement results in a solo LTR harboring TSDs. In contrast, a hallmark of the unequal inter-element recombination between the LTRs of two adjacent retrotransposons belonging to the same family is the lack of TSDs (Devos et al., 2002). We observed differences in the abundances of solo LTRs between the Elbe families (Figure 6b). As shown by BAC hybridization and analysis of the *B. vulgaris* genome assembly, there is a strong bias for Elbe3 solo LTRs. Factors influencing the recombination rate are LTR length and chromosomal localization (Du et al., 2010). However, Elbe LTRs are of similar lengths and similarly localized in subterminal chromosome regions. In *O. sativa*, evolutionarily young retrotransposons are characterized by a high percentage of intact elements and a low percentage of remnants and solo LTRs, but the opposite is found in older families (Ma et al., 2004). As Elbe3 elements are highly fragmented and many solo LTRs were found, we regard this family as the most ancient among Elbe retrotransposons. The high abundance of Elbe3 members is consistent with the strong Southern hybridization signals, even in the outgroup species *Spinacia oleracea* (Figure 7). However, based on the estimation of the integration time, the four Elbe families possess young and old members with transposition times ranging from 0.02 MYA (Elbe1) to 5.53 MYA (Elbe1-3). Furthermore, 22 of 27 Elbe members transposed at least 1 MYA, suggesting that this Errantivirus lineage is ancient and existed before species separation in the genus *Beta* (Table 1). Horizontal transfer cannot be excluded but is unlikely because there is no habitat overlap.

In *O. sativa* and *G. max*, the ratios of solo LTRs with and without TSDs are 1:3 and 1:10, respectively (Ma et al., 2004; Du et al., 2010). In *B. vulgaris*, the similar numbers of Elbe solo LTRs with TSDs and without TSDs reveal equal rates of intra-element and inter-element recombination, both process described by Devos et al. (2002).

The LTRs contain regions for transcription initiation (U3) and termination (U5). Therefore, the RNA transcript lacks the U3 at the 5′ end and the U5 at the 3′ end. During reverse transcription, full-length copies are generated and the missing regions are copied from the opposite LTRs through rearrangement at the RNA template. As multiple genomic RNAs are packaged during virus-like particle formation (Boeke et al., 1986; Jordan and McDonald, 1999; Feng et al., 2000), a template switch of the reverse transcriptase is very likely responsible for rearrangement of LTRs. Such reshuffling events have been described in *Saccharomyces cerevisiae*, *A. thaliana* and *G. max* (Jordan and McDonald, 1998; Marco and Marín, 2008; Du et al., 2009). In *B. vulgaris*, 56% of the detected Elbe solo LTRs show aberrant 5′ or 3′ ends, suggesting frequent sequence reshuffling events during reverse transcription (Figure 6b). Another explanation is recombination between the LTRs of different retrotransposon copies, resulting in circularization and deletion of the internal region and simultaneous production of a hybrid LTR containing the left end of the left LTR and the right end of the right LTR (Shirasu et al., 2000).

### The additional ORF of Elbe retrotransposons

As is typical for members of the *Errantiviridae*, Elbe retrotransposons encode an additional ORF, possibly originating from a nucleolus-derived gene encoding an Env-like protein. As described for membrane-penetrating proteins (Lerat and Capy, 1999), we found transmembrane domains in the *env*-like ORFs using two bioinformatic approaches; however, their number and position within the ORFs varied, indicating high sequence flexibility (Figure 3a).

Within the Envelope proteins of insect retrotransposons and mammalian retroviruses, the two conserved motifs KRG and LTPL have been reported to be important for the adsorption and penetration of a virus (Lerat and Capy, 1999). Due to the absence of the two motifs in the Sireivirus SIRE1 of *G. max* (Laten et al., 1998), it was assumed that Ty3-gypsy-like and Ty1-copia-like retrotransposons acquired the env ORFs independently (Kumar, 1998). In the plant Sireviruses Cotzilla of *B. vulgaris* and PpRT of *Pinus pinaster*, no conserved KRG and LTPL motifs were detected (Miguel et al., 2008; Weber et al., 2010). As neither the Env-like proteins of the Elbe elements nor the other Errantiviruses used for the alignment (Figure S3) contain the motifs, we assume that all plant retrotransposons lack these conserved motifs.
The Envelope proteins of G. max Errantiviruses and Sireviruses share common domains, indicating that they originated from a common ancestor. Therefore, independent capture of either one or both env-like genes after bifurcation of both superfamilies has been suggested (Du et al., 2010).

Implication of analysis of the Elbe lineage for Beta genome analysis

Detailed knowledge of the Elbe lineage, consisting of at least four families identified so far, is crucial for genome sequencing and annotation for two reasons. First, genome-wide distributed repetitive sequences of high abundance such as Elbe retrotransposons are difficult to assemble from the short sequence reads that are characteristic of high-throughput sequencing technologies. Second, annotation of truncated or internally deleted members of the Elbe families may be ambiguous: more than 50% of the identified Elbe retrotransposons (Figure 1) and 42% of the Elbe retrotransposons analyzed by BAC hybridization (Figure 6a) lack the reverse transcriptase gene that is a hallmark of retrotransposons. The Elbe3 family is the most abundant family, but none of its copies contain the gene encoding the reverse transcriptase (Figures 1 and 7c). These decayed members, fragments and solo-LTRs are difficult to identify; however, detailed knowledge of the Elbe retrotransposons and their numerous remnants enables straightforward annotation.

Sugar beet (B. vulgaris) is a relatively young crop originating from single crosses between mangold and fodder beet (Fischer, 1989), and hence has a narrow genetic base. Wild Beta species are an important resource in order to utilize the gene pool of the genus and introduce genetic diversity and agronomically desirable traits such as pathogen resistance and tolerance against abiotic stress (Van Geyt et al., 1990). Genome sequencing of wild beet species is in progress and will enable comparative studies of Elbe families, which in turn will support annotation of these genomes.

Elbe families are of considerable age, and their presence in various species of the genera Beta and Patellifolia suggest their existence in the last common ancestor. Interestingly, the env-like ORFs are different in each Elbe family, indicating independent acquisition and conservation over long evolutionary periods, but after the split of the genus Patellifolia, in which homologous env-like sequences are not detectable by Southern hybridization (Figure 7e–h). The function, if any, of the env-like ORF remains unknown. Despite the resemblance of transmembrane domains in the Env-domain, it is unlikely that Elbe retrotransposons represent intermediate stages in the evolution of retroviruses because the plant cell wall prevents infectivity. Nevertheless, the presence of the env ORFs shows that Elbe retrotransposons are vehicles for amplification and distribution of coding sequences within the genome. The Elbe env ORFs are most likely fragmented or truncated, but they provide sequences that may be assembled into new composite genes. Because of the ability to acquire genomic fragments including genes or parts thereof and to move them to new locations, env-like retrotransposons provide raw material for plant gene and genome evolution by increasing the gene copy numbers and changing the gene structure and regulation through transposition, insertion, excision and ectopic recombination (reviewed by Bennetzen, 2000).

Transposable elements, in particular LTR retrotransposons including Ty1-copia-like and Ty3-gypsy-like elements such as Elbe, are the most abundant sequence class of plant genomes. In plants with large and complex genomes, for example maize, transposable elements make up over 70% of the nuclear DNA (SanMiguel and Bennetzen, 1998). A consequence of the activity of retrotransposons is the dynamic change of the nuclear genome size. In addition to ancestral polyploidization, the activity of retrotransposons, often occurring in transpositional bursts, is the major cause for the variation of angiosperm genome size over several orders of magnitude (Bennett and Leitch, 1995, Bennetzen et al., 2005). Retrotransposons are a dynamic component of plant genomes, and analysis of the Elbe errantiviral lineage in B. vulgaris revealed different levels of retrotransposon rearrangement and reshuffling. Thus, the Elbe lineage provides insight into the evolution of highly abundant retrotransposon families exhibiting varying degrees of autonomy, age and fragmentation.

EXPERIMENTAL PROCEDURES

Plant material and isolation of genomic DNA

Plants of Beta vulgaris ssp. vulgaris (KWS2320), Betapatula (BETA 548), Beta corolliflora (BETA 846), Betana (BETA 541), Patellifolia procumbens (BETA 951) and Spinacia oleracea (Matador) were grown under long-day conditions. Wild beet seeds were obtained from the Plant Genome Resources Center Gatersleben (Germany), and cultivars were provided by KWS Saatzucht Einbeck (http://www.kws.de/).

Genomic DNA was isolated from young leaves using the CTAB (cetyltrimethyl/ammonium bromide) protocol as described previously (Saghai-Maroof et al., 1994).

Polymerase chain reaction

Probes for Elbe retrotransposons were amplified and sequenced for Southern hybridization. Primers used are listed in Table S1. PCR with 50 ng BAC DNA or plasmid template was performed in a 50 μl volume containing 5x GoTag buffer and 2.5 units of GoTag polymerase (Promega, http://www.promega.com). Standard PCR conditions were 94°C for 3 min, followed by 35 cycles of 94°C for 30 sec, the primer-specific annealing temperature for 30 sec, 72°C for 30–60 sec, and a final incubation at 72°C for 5 min.

BAC library analysis and BAC sub-cloning

To isolate a full-length Elbe retrotransposon, a high-density filter containing 9216 clones of a BAC library (Gindullis et al., 2001) was hybridized with an env-like RT sequence of B. vulgaris amplified using degenerated primer pairs (Vicient et al., 2001). DNA of

© 2012 The Authors
selected *B. vulgaris* BACs was isolated using the Nucleo Bond BAC 100 kit (Macherey & Nagel, http://www.mn-net.com/) and sub- 
com) using BamHI and HindIII. Sub-clones were hybridized with a 
radiolabeled env-like RT fragment. Positive clones were sequenced 
by primer walking using a Beckman Couter CEQ8000 capillary 
sequencer (Beckman Coulter, https://www.beckmancoulter.com).

**Southern hybridization**

Southern analyses were performed on HindIII-restricted genomic 
DNA separated on 1.2% agarose gels, which were transferred onto 
gehealthcare.com/). The membranes were hybridized with 32P-la- 
beled probes generated by random priming according to standard 
protocols at 60°C overnight (Sambrook et al., 1989). Subsequently, 
the filters were washed at 60°C in 2 × SSC/0.1% SDS and 1 × SSC/0.1% 
SDS for 10 min each, and signals were detected by autoradiogra- 
phy.

**Fluorescent in situ hybridization**

Mitotic chromosomes were prepared from the meristem of young 
plants. After incubation for 3–5 h in 2 mg 8-hydroxyquinoline, 
leaves were fixed in methanol:acetic acid (3:1). Fixed plant material 
was macerated in an enzyme mixture consisting of 0.3% w/v cyto- 

cellulase from *Aspergillus niger* (Sigma-Aldrich), 1.8% w/v 
cellulose from *A. niger* (Sigma-Aldrich), followed by spreading of the nuclei 
suspension onto slides as described by Desel et al. (2001).

**Fluorescent in situ hybridization** (FISH) of *B. vulgaris* chro- 
mosomes was performed as described by Heslop-Harrison et al. (1991) 
and modified for beet as described by Schmidt et al. (1994). FISH 
probes were labeled by PCR in the presence of biotin-11-dUTP. 
Standard stringencies of 76% for hybridization and 79% for washing 
were used. Chromosome preparations were counter-stained using 
DAPI (4',6'-diamidino-2-phenylindole) and mounted in antifade 
solution (CitiFluor, http://www.citifluor.co.uk). Examination of 
slides was performed using a fluorescent microscope (Zeiss Axi- 
oplan 2 imaging Zeiss, http://www.zeiss.com) equipped with filter 09 
(fluorescein isothiocyanate), filter 15 (Cy3) and filter 02 (DAPI). 
Images were acquired directly using APEX SPECTRAL IMAGING 
spectral-imaging.com) coupled with the high-resolution CCD 
camera ASI BV300-20A.

**Computational methods**

In order to identify complete Elbe copies, homology searches were 
performed using *B. vulgaris* sequence entries in the EMBL data- 
base, mainly consisting of BAC end sequences retrieved by 
McGrath et al. (2004). Retrieved sequences were used to assemble an 
in silico env-like Ty3-gypsy retrotransposon from which primers 
were derived.

Using the BLAST algorithm (Altschul et al. 1990), a preliminary 
*B. vulgaris* genome assembly was queried using family-specific 
LTRs to identify additional Elbe members. This unedited and non- 
public assembly comprises 768.1 Mb in 383,968 contigs, with an 
N50 contig size of 5399 bp. A current version of the sugar beet 
genome draft (RefBeet0.9) is available for download at http:// 
vbsq.molgen.mpg.de (Lange et al., 2010; Dohm et al., 2012).

Hits of at least 1400 bp and more than 80% identity to the queries 
were further analyzed. Sequences were aligned using the MUSCLE 
algorithm (Edgar, 2004). Comparative sequence analyses were 
performed using MEGA 4.0 software (Tamura et al., 2007). Neigh- 
bor-joining trees were constructed with 1000 bootstrap replicates. 
Evolutionary distances were computed using the Poisson correction 
method, and positions containing alignment gaps and missing data 
were eliminated only in pairwise sequence comparisons. Errantivi- 
rus transposition times were estimated using the equation \( t = K/2r \), 
where \( t \) is the age, \( K \) is the number of nucleotide substitutions per 
site between each LTR pair, and \( r \) is the nucleotide substitution rate. 
A mean synonymous substitution rate of 1.5 × 10⁻⁸ mutations per 
site per year as determined for the chalcone synthase and Adh loci 
in *A. thaliana* (Koch et al., 2000) was applied.

Protein sequences of Elbe members were deduced using the 
GeneWise algorithm (Birney et al., 2004) with Cyclops-2 (Chavanne 
et al., 1998) as the reference sequence.

The Plant Care database (http://bioinformatics.psb.ugent.be/ 
etal) was used for promoter motif scanning within the LTRs.

To predict the putative transmembrane (TM) domains in the env-like 
ORFs, TMPred (http://www.ch.embnet.org/software/TMPRED_form. 
html) was used for promoter motif scanning within the LTRs.

**Retrotransposon accessions**

The accession numbers of the retrotransposons analyzed in this 
paper are listed in Table S2.

**ACKNOWLEDGMENTS**

We thank Ines Walter for excellent technical assistance. This work 
was funded by the Verbundprojekt GABI BeetSeq: Erstellung einer 
Referenzsequenz für das Genom der Zuckerrübe (*Beta vulgaris*), 
FKZ 0315069A and 0315089B (to H.H. and B.W.), and by the German 
Federal Ministry of Education and Research (KMU Innovativ, grant 
number 0315425).

**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online 
version of this article:

**Figure S1.** Overview of promoter motifs and putative polyadenyla- 
tion signals within the LTR.

**Figure S2.** Amino acid sequence signatures of the gag-pol polypro- 
tein.

**Figure S3.** Alignment of the putative Env-like protein.

**Figure S4.** Prediction of putative TM domains in the Env amino acid 
sequences of representative members of all Elbe families (examples 
are also shown in Figure 3).

**Table S1.** Primers used for amplification of domain-specific probes 
derived from the consensus sequence.

**Table S2.** Accession numbers of Elbe retrotransposons and other 
sequences used as references. Please note: As a service to our authors and readers, this journal 
provides supporting information supplied by the authors. Such 
materials are peer-reviewed and may be re-organized for online 
delivery, but are not copy-edited or typeset. Technical support 
issues arising from supporting information (other than missing 
files) should be addressed to the authors.

**REFERENCES**


Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic 

Arumuganathan, K. and Earle, E.D. (1994) Nuclear DNA content of some 


