Whole genome sequence comparison of vtx2-converting phages from Enterotoaggregative Haemorrhagic Escherichia coli strains

Grande et al.
Whole genome sequence comparison of\nvtx2-converting phages from Enteroaggregative\nHaemorrhagic \textit{Escherichia coli} strains

Laura Grande$^{1,2,*}$, Valeria Michelacci$^1$, Rosangela Tozzoli$^1$, Paola Ranieri$^1$, Antonella Maugliani$^1$, Alfredo Caprioli$^1$ and Stefano Morabito$^1$

Abstract

\textbf{Background:} Enteroaggregative Haemorrhagic \textit{E. coli} (EAHEC) is a new pathogenic group of \textit{E. coli} characterized by the presence of a \textit{vtx2}-phage integrated in the genomic backbone of Enteroaggregative \textit{E. coli} (EAggEC). So far, four distinct EAHEC serotypes have been described that caused, beside the large outbreak of infection occurred in Germany in 2011, a small outbreak and six sporadic cases of HUS in the time span 1992–2012. In the present work we determined the whole genome sequence of the \textit{vtx2}-phage, termed Phi-191, present in the first described EAHEC O111:H2 isolated in France in 1992 and compared it with those of the \textit{vtx}-phages whose sequences were available.

\textbf{Results:} The whole genome sequence of the Phi-191 phage was identical to that of the \textit{vtx2}-phage P13374 present in the EAHEC O104:H4 strain isolated during the German outbreak 20 years later. Moreover, it was also almost identical to those of the other \textit{vtx2}-phages of EAHEC O104:H4 strains described so far. Conversely, the Phi-191 phage appeared to be different from the \textit{vtx2}-phage carried by the EAHEC O111:H21 isolated in the Northern Ireland in 2012. The comparison of the \textit{vtx2}-phages sequences from EAHEC strains with those from the \textit{vtx}-phages of typical Verocytotoxin-producing \textit{E. coli} strains showed the presence of a 900 bp sequence uniquely associated with EAHEC phages and encoding a tail fiber.

\textbf{Conclusions:} At least two different \textit{vtx2}-phages, both characterized by the presence of a peculiar tail fiber-coding gene, intervened in the emergence of EAHEC. The finding of an identical \textit{vtx2}-phage in two EAggEC strains isolated after 20 years in spite of the high variability described for \textit{vtx}-phages is unexpected and suggests that such \textit{vtx2}-phages are kept under a strong selective pressure. The observation that different EAHEC infections have been traced back to countries where EAggEC infections are endemic and the treatment of human sewage is often ineffective suggests that such countries may represent the cradle for the emergence of the EAHEC pathotype. In these regions, EAggEC of human origin can extensively contaminate the environment where they can meet free \textit{vtx}-phages likely spread by ruminants excreta.

\textbf{Keywords:} Enteroaggregative haemorrhagic \textit{E. coli}, \textit{vtx}-phages, Whole genome sequence, Tail fibers

* Correspondence: laur.grande@gmail.com
$^1$EU Reference Laboratory for \textit{E. coli}, Veterinary Public Health and Food Safety Department, Istituto Superiore di Sanità, Viale Regina Elena 299, Rome 00161, Italy
$^2$Università degli Studi di Roma Tre, Rome, Italy

© 2014 Grande et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

Diarrheagenic Escherichia coli (DEC) are a heterogeneous group of pathogenic E. coli causing a wide range of enteric diseases in humans and animals [1].

Enteroaggregative E. coli (EAggEC) are a DEC pathotype inducing a gastrointestinal illness characterized by long-lasting watery, mucoid, secretory diarrhoea with low-grade fever and little or no vomiting [2,3]. EAggEC infections are a common cause of acute diarrheal illness among children in low-income countries, but sporadic cases and outbreaks are recorded in industrialized countries as well [4,5].

In 2011, a large E. coli outbreak struck Germany causing more than 4,000 human infections including 50 deaths [6]. The outbreak strain, an E. coli O104:H4, showed the presence of the typical virulence genes of EAggEC such as aggR, aaiC, sepA, aatA and, at the same time, it carried a bacteriophage conveying the genes encoding the Verocytotoxin (Vtx) subtype 2a (vtx2a) [7]. In accordance with this genomic asset, the strain showed the Enterogaugregative typical “stacked brick” adhesion to cultured Hep-2 cells and was able to produce Vtx2 [8]. The infecting strain thus displayed an unusual combination of virulence features comprising the colonization repertoire from EAggEC coupled with the production of a toxin typically produced by Vtx-producing E. coli (VTEC), a DEC type causing haemorrhagic colitis and Haemolytic Uremic Syndrome (HUS) worldwide [1].

The impact of the German outbreak was so huge that the epidemic strain became iconic of a new DEC type: the Enteroaggregative Haemorrhagic E. coli (EAHEC) [9]. The occurrence of the German outbreak also caused the scientific community to look retrospectively at the reported HUS cases linked to infections with atypical VTEC types or to browse the scientific literature in order to assess if other EAHEC cases of infection could be retrieved. It turned out that in the time period 1992–2012 a small outbreak and at least six sporadic cases of HUS had been described as being associated with EAHEC strains belonging to four different serotypes: O111:H2, O86: HNM, O104:H4 and O111:H21 [8,10-13].

The analysis of the whole genome sequence of the EAHEC O104:H4 that caused the German outbreak in 2011 showed that the vtx2-phage is inserted in a bacterial genomic backbone typical of EAggEC [14], therefore the EAHEC pathotype seems to have arisen from the acquisition of vtx2-phages by classical EAggEC strains.

The appearance of the EAHEC group has shown that the stable acquisition of vtx-phages seems to have occurred at least twice by two different DEC groups, the EAggEC and the atypical EPEC (aEPEC) from which the typical VTEC pathotype derives [15-17]. Moreover, the ability of vtx2-phages to infect, in the laboratory conditions, different E. coli pathogroups including ExPEC has been reported [18,19]. This observation, together with the isolation of Enterobacteriaceae other than E. coli producing Vtx from cases of human disease [20,21] suggests that vtx-phages can infect a range of bacterial hosts wider than expected, confirming the pivotal role of phages in the evolution of bacterial pathogens.

In the present work we determined the whole genome sequence of the vtx2-phage present in the first EAHEC ever described and compared it with that of the vtx2-phages present in the EAHEC O104:H4 and O111:H21 available in the public repositories and with those of other vtx-phages, with the aim of investigating the mechanisms underlying the evolution of the EAHEC pathotype.

Methods

Bacterial strains

The EAHEC O111:H2 strain ED 191 has been used to obtain the vtx2-phage subjected to whole genome sequencing and is part of the collections held at Istituto Superiore di Sanità. The strain’s characteristics have been described in a previous publication [10].

E. coli K12 strain LE392 [22] has been used as a propagator strain in infection experiments for the vtx2-phage amplification prior to sequencing.

Determination of the vtx2-phages integration sites in the E. coli genome

The vtx2-phage integration site in the E. coli strain ED 191 has been determined. The occupancy of loci sbcB, wrbA, yehV, Z2577, and yecE has been assessed as previously described [18].

Infection experiments and phages propagation

The EAHEC strain ED 191 has been exposed to UV light in order to induce the excision of phage genome from the bacterial chromosome [23]. In detail, the bacterial strain has been grown in Luria-Bertani (LB) broth (Oxoid Limited, Basingstoke Hampshire, UK) overnight at 37°C with vigorous shaking. The culture has been diluted 1:100 in LB modified broth (LB with 0.001% thiamine V/V) and grown to 0.5 OD 600, pelleted and re-suspended in a sterile solution of CaCl2 10 mM. The culture has been centrifuged and the supernatant containing bacteriophages present in the EAHEC O104:H4 and O111:H21 [8,10-13].

The analysis of the whole genome sequence of the EAHEC O104:H4 that caused the German outbreak in 2011 showed that the vtx2-phage is inserted in a bacterial genomic backbone typical of EAggEC [14], therefore the EAHEC pathotype seems to have arisen from the acquisition of vtx2-phages by classical EAggEC strains.

The appearance of the EAHEC group has shown that the stable acquisition of vtx-phages seems to have occurred at least twice by two different DEC groups, the EAggEC and the atypical EPEC (aEPEC) from which the typical VTEC pathotype derives [15-17]. Moreover, the ability of vtx2-phages to infect, in the laboratory conditions, different E. coli pathogroups including ExPEC has been reported [18,19]. This observation, together with the isolation of Enterobacteriaceae other than E. coli producing Vtx from cases of human disease [20,21] suggests that vtx-phages can infect a range of bacterial hosts wider than expected, confirming the pivotal role of phages in the evolution of bacterial pathogens.

In the present work we determined the whole genome sequence of the vtx2-phage present in the first EAHEC ever described and compared it with that of the vtx2-phages present in the EAHEC O104:H4 and O111:H21 available in the public repositories and with those of other vtx-phages, with the aim of investigating the mechanisms underlying the evolution of the EAHEC pathotype.

Methods

Bacterial strains

The EAHEC O111:H2 strain ED 191 has been used to obtain the vtx2-phage subjected to whole genome sequencing and is part of the collections held at Istituto Superiore di Sanità. The strain’s characteristics have been described in a previous publication [10].

E. coli K12 strain LE392 [22] has been used as a propagator strain in infection experiments for the vtx2-phage amplification prior to sequencing.

Determination of the vtx2-phages integration sites in the E. coli genome

The vtx2-phage integration site in the E. coli strain ED 191 has been determined. The occupancy of loci sbcB, wrbA, yehV, Z2577, and yecE has been assessed as previously described [18].

Infection experiments and phages propagation

The EAHEC strain ED 191 has been exposed to UV light in order to induce the excision of phage genome from the bacterial chromosome [23]. In detail, the bacterial strain has been grown in Luria-Bertani (LB) broth (Oxoid Limited, Basingstoke Hampshire, UK) overnight at 37°C with vigorous shaking. The culture has been diluted 1:100 in LB modified broth (LB with 0.001% thiamine V/V) and grown to 0.5 OD 600, pelleted and re-suspended in a sterile solution of CaCl2 10 mM. The culture has been exposed to UV light (130 μJoule X 100) in a crosslinker “Stratalinker” UV crosslinker (Stratagene Cloning Systems, La Jolla, CA, USA). After induction, the culture has been diluted in LB modified broth and incubated at 37°C for 5 hours with vigorous shaking. The culture has been centrifuged and the supernatant containing phages particles filtered with 0,22 μm pore-filters. 100 μl of phage particles suspension have been added to 100 μl of a culture of the propagator strain E. coli LE392 grown in LB modified broth at 0.5 OD 600 and maintained at 37°C for 20 minutes with static incubation. Each tube has been added with 3.5 ml of LB modified soft agar (LB modified broth with agar 7 g/L) at 42°C and immediately
poured on LB modified agar plates (LB modified broth with 15 g/L agar). Plates have been incubated overnight at 37°C.

Four ml of SM buffer (100 mM NaCl, 8 mM MgSO₄·7H₂O, 50 mM Tris–HCl 1 M pH 7.5, Gelatin 0.002%) have been dispensed to each plate in order to recover phages particles from the lytic plaque and kept overnight at 4°C. The phage suspension in SM has been recovered and chloroform has been added at 5% final concentration. The phage suspension has been centrifuged at 500g 10 minutes twice for removing agar debris and used to re-infect the propagator E. coli strain LE392 in the conditions described above in order to increase the phage titre. Finally, the phage suspension has been concentrated by using Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-30 tubes (Merck Millipore, Billerica, MA, USA) with a cut-off of 30 KDa. Final phage titre was 7 × 10¹⁰ PFU/ml.

CsCl gradient and viral DNA extraction

The suspension has been purified by Isopycnic Centrifugation through CsCl Equilibrium gradient as described by Sambrook and Russell [23]. Briefly, 2 ml of the phage suspension have been added with 1.5 g of CsCl transferred to ultracentrifuge tubes, which have been filled with a CsCl solution 0.75 g/ml. The tubes have been finally sealed with mineral oil and centrifuged in a Beckman ultracentrifuge at 154,000xg, 8°C for 20 hours in a SW-41 rotor. The band containing the phage particles has been collected with a syringe by puncturing the tube. The recovered solution containing the purified phage particles has been dialyzed in against 10 mM NaCl, 50 mM Tris–HCl pH 8.0, 10 mM MgCl₂. A final volume of 1 ml was obtained.

The suspension has been treated by adding 100 units of DNase I RNase-free (New England Biolabs, USA) at 37°C for one hour to eliminate free DNA contaminating the phage suspension. Finally a treatment with proteinase K 50 μg/ml at 56°C for one hour has been carried out to disrupt the phage capsid followed by DNA extraction with phenol-chloroform-isoamyl alcohol [23]. Phage DNA concentration after the purification step was estimated to be 239.4 ng/μl.

Library preparation and whole genome sequencing of the phage DNA

Phage DNA has been sequenced with an Ion Torrent PGM semiconductor sequencer (Life Technologies, Carlsbad, USA) using the 200 bp protocol. An Ion Torrent 314 chip has been used following the manufacturer instructions (Life Technologies, Carlsbad, USA). Genomic library has been used following the manufacturer instructions (Life Technologies, Carlsbad, USA). Using the Ion OneTouch 200 Template kit and instruments (Life Technologies, Carlsbad, USA).

Assembly and further bioinformatics analyses

The reads resulting from the sequencing of the vtx2-phage DNA from the EAHEC O111:H2 strain, termed Phi-191, have been assembled in contigs by using the open source MIRA software integrated in the Ion Torrent Server. Contigs have been imported in Kodon software (Applied Maths NV, Sint-Martens-Latem, BE) for analysis. To fill in the gaps between contigs, a total of 95 primers have been designed and used for sequencing by Sanger technology using a Genetic analyzer 3130 (Life Technologies, Carlsbad, USA). Mauve software [24] has been used to order the contigs using the sequence of the phage P13374 from the E. coli O104:H4 that caused the outbreak in Germany in 2011 [GenBank: NC_018846.1] as reference. The complete sequence of the Phi-191 phage has been annotated by Prokka tool on the online server Galaxy/CRS4 [25] and submitted to GenBank [GenBank: KP971864]. The G + C content has been analysed by the GC calculator free online tool [26]. Identification of putative tRNA genes has been performed using tRNAscan-SE [27]

The raw sequence data (short reads) from the EAHEC O111:H21 strain 226 were retrieved from the SRA database present on NCBI website [NCBI SRA: SRA055981] and aligned on the complete sequence of Phi-191 phage, determined in the present study and used as reference, with the Bowtie2 free software implemented in the Galaxy/CRS4 server [25].

Genomic comparisons between the available vtx-phages sequences have been performed by using the BLAST algorithm available at NCBI [28] and the Mauve free alignment software [24]. Comparison map between vtx-phages has been generated by Circoletto online tool [29,30]. For the pictogram construction, bit-score values have been used to describe the quality of the alignment at a given point. The bit-score is a normalized version of the score value returned by the BLAST searches, expressed in bits [28].

Ethics

This work does not include animal testing and does not report human data. All the information regarding outbreaks and cases of infection are all from already published papers properly referenced in the text.

Results

Sequencing of the vtx2-phage from the EAHEC O111:H2

The whole genome sequencing of the vtx2-phage from the EAHEC O111:H2 strain, Phi-191, produced 320,044 reads of a mean length of 204 bp, for a total of 65.30 Mb sequenced. The assembly of the Phi-191 sequence reads
using as a reference the genome of the P13374 phage [GenBank: NC_018846.1] produced 151 contigs that were further analysed using bioinformatics resources. The estimated coverage of the phage genome was 1088X.

The length of the whole sequence of Phi-191 resulted to be about 61 Kb (61,036 bp) [GenBank: KF971864] with a mean G + C content of 50.2%.

Sequence annotation revealed the presence of 87 predicted coding sequences including the genes encoding the subtype 2a of the Verocytotoxin and three transfer RNAs (tRNAs) (Figure 1).

As it has been reported for the P13374 phage, conveying the vtx2 genes in the EAHEC O104:H4 strain that caused the German outbreak in 2011 [31], the genome of Phi-191 (Figure 1) only included the lambda genes cl and cro while the other genes typically composing the regulatory repertoire of lambda phages such as cII, cIII, N, Ea10 and gam seemed not to be present [31]. Most of these genes are involved in the regulation of the switch between lysogeny and lytic cycle in lambda-like phages. Therefore, their absence suggests the existence of an alternative mechanism used to regulate the choice of entry into lysogenic state. The Phi-191 phage is integrated in the bacterial chromosome in the wrbA locus, as it has been described for P13374 [31].

**Comparison of the Phi-191 whole genome sequence with other vtx-phage sequences**

A BLAST search of the Phi-191 vtx2-phage whole genome sequence against those collected in the nucleotide repository held at NCBI, returned a 99% identity with that of the P13374 phage from O104:H4 strain CB13374 which caused the German outbreak in 2011 [GenBank: NC_018846.1] and a 91% identity with that of the vtx2-phage TL-2011c carried by the VTEC O103:H25 strain that caused a severe HUS outbreak in Norway in 2006 [GenBank:JQ011318] [32]. This finding is in agreement with how reported in the literature for the vtx-phage carried by the EAHEC O104:H4 from the German outbreak and the TL-2011c phage [33]. As expected, the Phi-191...
The phage sequence also showed 99% identity with the vtx2-phage from another EAHEC O104:H4 strain isolated during the 2011 German outbreak, the strain 2011C-3493 [GenBank:CP0032891.1] and was highly related (97% nucleotide sequence identity) to the sequence of two vtx2-phages from two EAHEC O104:H4 strains isolated from as many haemorrhagic colitis cases occurred in Georgia in 2009 [GenBank:CP003301.1, CP003297.1] [34].

The other scores included hits with query coverage values ranging from 87% down to 60% and having 98%-99% sequence identity, with a number of other vtx-phages identified in different VTEC strains (Table 1). The alignment of all the phage sequences comprised in the 100%-60% similarity range is shown in a pictogram generated with the Circletto online tool [30] (Figure 2) and in another one produced by Mauve software (in Additional file 1: Figure S1).

Recently, another EAHEC strain of serotype O111:H21 (strain 226), isolated from an HUS case occurred in Northern Ireland in 2012, has been described [13]. The short reads from the whole genome-sequencing project of this strain are available in NCBI sequence reads archive [NCBI SRA: SRA055981] and have been used for comparison. Unfortunately, among the 456 contigs obtained from the de novo assembly of the reads, the one including the vtx2 genes was only 8,042 bp long, hindering the analysis of the complete genomic sequence from the Phi-191 phage. Nevertheless, its whole-genome sequencing reads were aligned to the complete genomic sequence from the Phi-191 phage. The alignment was carried out with Bowtie2 software and visualized on the Integrative Genomics Viewer (IGV) free tool [45] (data not shown). This analysis failed to identify, in the genome of strain 226, a vtx2-phage with characteristics similar to that of Phi-191. In particular, the alignment showed that the reads from strain 226 mapped on the sequence of the Phi-191 phage only in the region comprising the Vtx2 A subunit-coding gene and extending 4.2 kb downstream the gene encoding the Vtx2 B subunit. It should be noted that alignments conducted using the whole short reads set could have generated a too high background noise deriving from the presence of sequences from other bacteriophages integrated in different regions of the chromosome of the O111:H21 strain 226. Nevertheless, the lack of sequencing reads aligning to the rest of the sequence of Phi-191 suggests that a different vtx2 phage intervened in the emergence of this last EAHEC strain.

| Strain                          | Similarity% | Acc. No. | Reference |
|--------------------------------|-------------|----------|-----------|
| E. coli O145:H28 str. RM13514   | 87%         | CP006027.1 | [35]      |
| E. coli O103:H2 str. 12009 DNA  | 86%         | AP010958.1 | [35,36]   |
| Phage VT2 phi_272              | 85%         | HQ424691.1 | -         |
| E. coli O157:H7 str. TW14359    | 67%         | CP001368.1 | [37]      |
| E. coli O157:H7 str. EC4115     | 67%         | CP001164.1 | [38]      |
| E. coli O111:H str. 11128       | 65%         | AP010960.1 | [36]      |
| E. coli O157:H7 str. Sakai      | 65%         | BA000007.2 | [39]      |
| E. coli Xuzhou21                | 65%         | CP001925.1 | [40]      |
| E. coli O157:H7 EDL933          | 65%         | AE005174.2 | [41]      |
| Stx2 converting phage II        | 64%         | AP005154.1 | [42]      |
| Stx2 converting phage I         | 63%         | AP004402.1 | [43]      |
| Enterobacteria phage Min27      | 63%         | EU311208.1 | [44]      |
| Stx1 converting phage           | 60%         | AP005153.1 | [42]      |

The similarity score indicates the query coverage values with 98%-99% sequence identity. The hits with similarity values down to 60% are shown.

Genomic comparison of Phi-191 with other vtx2-phages
In order to delve into the genomic organization of the different vtx2-phages we performed a progressive Mauve alignment between all the phages sequences used in the nucleotide comparison (Additional file 1: Figure S1). This analysis confirmed the presence of similarities between the sequences but also highlighted that an extensive rearrangement between the sequence-blocks must have occurred at some points during the phages’ evolution. This observation is in agreement with how reported in literature [46-48].

A deeper analysis of the Phi-191 genome was conducted by using phage sequences selected among those from typical VTEC displaying the highest similarity values (Table 1). The alignment of the Phi-191 sequence with those of phage TL-2011c (O103:H25), and the phages from the VTEC strains RM13514 (O145:H28) and 12009 (O103:H2), the causative agent of a romaine lettuce-associated outbreak occurred in the US in 2010 [35] and isolated in Japan from a sporadic case of diarrhoea [36], respectively, showed the presence of two sequence blocks that seemed to be peculiar to the EAHEC vtx2-phage and divergent from the same regions in the three classical VTEC associated-phages (Figure 3). One of the two regions was 1,500 bp long and comprised 140 bp of the 5’-terminus of a gene encoding a lysozyme (rrrd), two complete genes encoding the lysis protein S and a hypothetical protein, and 440 bp of the 3’-terminus of a gene annotated as a hypothetical protein. The 1,500 bp region corresponded to the nucleotide positions 5,240-6,725 in the P13374 genome (ORFs 12–15). The other fragment was 900 bp long and spanned the region comprised between nucleotides 42,160 and 43,050 in the genome of the P13374 phage. The 900 bp fragment shared 100% homology with a region of 730 bp at the 3’-terminus of the ORF65 of P13374 phage encoding a phage tail fiber and a 72 bp fragment of the ORF66 coding for a tail fiber adesine.
Distribution and analysis of the putative EAHEC-associated sequence regions amongst the vtx-phages

A BLAST comparison was conducted with the aim of investigating the distribution of the two putative EAHEC-associated regions among all the vtx-phage sequences available at NCBI. Such an analysis showed that only the 900 bp region encoding phage fiber tail was present in all the fully sequenced vtx2-phages from EAHEC strains, [GenBank:CP0032891.1; CP003301.1; CP003297.1] and divergent or absent in all the other phage sequences investigated. Accordingly, the presence of the sole sequence of 900 bp was assessed in the contigs derived from the de novo assembly of the short reads from the genome sequence of the O111:H21 strain 226 from Northern Ireland.

Discussion

The E. coli genome continuously changes through both small-scale variations and horizontal gene transfer of mobile genetic elements (MGE). Among MGEs, bacteriophages play a pivotal role in the evolution of E. coli pathogenic clones [49] by providing a mean for genomic remodelling and conveying important virulence genes such as those encoding the Verocytotoxins (vtx1 and vtx2) in the Verocytotoxin-producing E. coli (VTEC).

In 2011 a huge outbreak caused by an Enteroaggregative Haemorrhagic E. coli (EAHEC) O104:H4 struck Germany with more than 3,000 cases of infection, 800 HUS, and 50 deaths [12]. The causative agent was a mosaic strain deriving from the lysogenization of an Enteroaggregative E. coli strain with a vtx2-phage [8]. Such a virulence combination was indeed associated with elevate pathogenicity, as demonstrated by the high rate of human infections evolving to HUS, even among adults (88% and 42 years of median age), and the heavy toll of 50 deaths [6].

This arrangement of virulence factors in E. coli strains from human disease had been occasionally reported before during the investigation of a small outbreak of HUS...
occurred in France in 1992 and a case of infection in Japan in 1999 [10,11].

The occurrence of the German outbreak caused the scientific community to look back retrospectively at the repositories of VTEC infections records and culture collections and it turned out that some other sporadic cases of infections with Vtx2-producing Enterogaegregative \textit{E. coli} O104:H4 had already occurred in Europe and Asia in the time span 2001–2011 [12,31]. Finally a HUS case occurred in Northern Ireland in 2012 was demonstrated to be associated with an EAHEC O111:H21 [13].

The observation of sporadic cases and outbreaks occurring throughout a 20-years time span, all caused by Vtx2-producing EAggEC and belonging to different serotypes, strengthens the hypothesis that these pathogenic \textit{E. coli} represent a new pathogroup, as it has been recently proposed [9].

To better understand the events underlying the emergence of EAHEC we determined the whole genome sequence of Phi-191, the \textit{vtx2}-phage present in the EAHEC O111:H2 isolated during the French outbreak of 1992, and compared it with the sequences of the \textit{vtx2}-phage present in the EAHEC strains described in the following years and available in GenBank.

Interestingly, the genomic sequence of Phi-191 was almost identical to that of the \textit{vtx2}-phages from the EAHEC O104:H4 strains isolated during the 2011 German outbreak about 20 years later. This is noteworthy since \textit{vtx}-phages are characterized by a high degree of variability [50,51]. It is conceivable that the same \textit{vtx2}-phage has been acquired in two different events and that the selective pressure impeded the accumulation of changes in the phage sequence before the phage infection events occurred.

However, the EAHEC O111:H21 isolated in Northern Ireland in 2012 seems to host a different type of \textit{vtx2}-phage, suggesting that at least two different \textit{vtx2}-phage types have been successfully transferred to EAggEC. Unfortunately, the sequence of the phage of the EAHEC O86: HNM isolated in Japan in 1999 was not available for comparison [11].

It has been hypothesized that the infection with a lambdoid phage can be mediated by the cross-talking between the bacterium and the phage resulting in host specificity [52]. An extended comparison of the EAHEC \textit{vtx2}-phages with the whole genome sequences of \textit{vtx}-phages from VTEC strains available at NCBI returned a wide range of similarities between sequences, going from 87% to 60% and lower. This picture is in line with how reported for the general variability of \textit{vtx}-phages sequences [50]. Interestingly, one region of 900 bp, identified in the Phi-191 and encoding a tail fiber, was present in all the \textit{vtx2}-phages from EAHEC and was also present in the short reads dataset from the EAHEC O111:H21. At the same time this DNA sequence was
absent in all the vtx-phage sequences identified in VTEC strains and stored at NCBI.

This is in agreement with previously reported data which pointed at a larger fragment including this region as one of those differentiating the P13374 genome from the E. coli phage TL-2011c (O103:H25) and not exhibiting significant homology to known vtx-encoding phages [31].

The annotation of the Phi-191 genome showed that this sequence peculiar to EAHEC vtx2-phages contains part of a gene encoding a type of phage tail fiber displaying some conserved aminoacidic motifs such as a Collagen triple helix repeat (20 copies) [NCBI CDD:189968] and the Peptidase_S74 [NCBI CDD:258151]. The latter is the C-terminal domain of the bacteriophage protein endosialidase, which forms homotrimmeric molecules and releases itself from the end-tail-spike of the bacteriophages [53].

The 900 bp-long sequence could potentially encode part of the mechanism defining the specificity of the vtx2-phages for EAggEC strains, being directly involved in the phage-bacterium interaction. As a matter of fact, several authors reported that the interactions between phage tail fibers and host proteins, such as LamB and OmpC [52,53] contribute to the success of the infection, as demonstrated by the finding that lamB gene mutations block phage adsorption [52]. It is therefore conceivable that differences in phage tail fibers may contribute to define vtx2-phages tropism for E. coli recipients, although this hypothesis together with the mechanisms underlying this process still need to be verified.

For a successful infection to occur, suitable vtx-phages and E. coli acceptors need to meet in the same environment. In the case of the emergence of typical VTEC, the events of vtx-phage acquisition probably occurred at the level of the gastrointestinal tract of ruminants [54] where both vtx-phages and eAEPEC are abundant [55,56].

Conversely, the EAHEC emergence is probably not directly connected to an animal reservoir since EAggEC are human pathogens with an inter-human transmission of the infection [1]. The environment, in turn, plays a role in the pathogen’s amplification cycles, particularly in geographic areas characterised by poor hygienic conditions, where the lack of effective human sewage treatments make the infections with enteric pathogens, including EAggEC, endemic [57]. In such a scenario, an environment contaminated with ruminant’s excreta might have been the source of the vtx2-phages found in EAHEC as it has been recently proposed [16]. Such a picture may account for the existence of a favourable setting for the EAggEC and the vtx-phages to come in contact and for the following selection process resulting in the occasional emergence of an E. coli strain matching the EAHEC definition.

Conclusions
The vtx2-phages characterising EAHEC seem to belong to a sub-population of vtx-phages kept under selective pressure and characterised by the presence of a gene encoding a tail fiber, which could be involved in the mechanism used to recognize the EAggEC. The new EAHEC pathogroup may have emerged following multiple vtx2-phage acquisition events favoured by an overlapping of a human reservoir of pathogenic E. coli, the EAggEC, with the known animal reservoir of vtx-phages. The emergence of this new E. coli pathogroup further witnesses the great adaptability and plasticity of this bacterial species and underlines the need to rethink the global asset of hygienic practices to mitigate enteric infections worldwide; particularly in the presence of a global market of food-stuffs that is extending its boundaries towards low-income countries in the quest of new sources to meet the always increasing demand of cheap and exotic food commodities.

Additional file

Additional file 1: Figure S1. Mauve Progressive Alignment of Phi-191 genome with vtx-phage genomes from EAHEC and VTEC strains showing the highest score of similarity. Blocks with the same colours indicate the vtx-phages regions with identical DNA sequence. White fragments in a phage sequence indicate regions lacking of correspondence in the other sequences. Connecting lines link the same genomic block in different genomes and help to pinpoint the re-arrangement between vtx-phage genomes.

Abbreviations
aEPEC: Atypical enteropathogenic Escherichia coli; DEC: Diarrheagenic Escherichia coli; EAggEC: Enteroaggregative Escherichia coli; EAHEC: Enteroaggregative haemorrhagic Escherichia coli; HC: Haemorrhagic colitis; HUS: Haemolytic uremic syndrome; MGE: Mobile genetic elements; VTEC: Verocytotoxin-producing Escherichia coli; vtx: Verocytotoxin genes.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
GL carried out the infection experiments, contributed to the sequencing and revised critically the manuscript; RP carried out part of the whole genome sequencing and bioinformatics analysis; TR contributed to all the steps of the whole genome sequencing and revised critically the manuscript; RP carried out part of the infection experiments and CilC gradient; MA helped in the revision of the draft manuscript; CA contributed to the revision of the draft manuscript; MS conceived the study and strongly contributed to revise the manuscript. Finally, all the authors approved the manuscript to be published.

Acknowledgements
We thank Prof. Muniesa for hosting the corresponding author at University of Barcelona, Spain, where the basic techniques for working with phages have been acquired. We also thank Prof. Iniguez and Dr. Muniesa for their support to this work.

Received: 15 April 2014 Accepted: 3 July 2014
Published: 8 July 2014

References
1. Nataro JP, Kaper JB. Diarrheagenic Escherichia coli. Clin Microbiol Rev 1998, 11(1):142–201.
2. Bhan MK, Khoshoo V, Sommerfelt H, Raj P, Sazawal S, Srivastava R. Enteroaggregative Escherichia coli and Salmonella associated with nontyphoid persistent diarrhea. Pediatr Infect Dis J 1989, 8(8):499–502.
3. Paul M, Tsukamoto T, Ghosh AR, Bhattacharya SK, Manna B, Chakrabarti S, Nair GB, Jack DA, Sen D, Takeda Y: The significance of enterogastronereactive Escherichia coli in the etiology of hospitalized diarrhoea in Calcutta, India and the demonstration of a new honey-combed pattern of aggregative adherence. *FEMS Microbiol Lett* 1994, 117(3):319–325.

4. Huang DB, Nataro JP, DuPont HL, Kamat PP, Mistretta AD, Oluwaseyi PC, Chiang T: Enterogastronereactive Escherichia coli is a cause of acute diarrheal illness: a meta-analysis. *Clin Infect Dis* 2006, 43(5):556–563.

5. Estrada-García T, Navarro-García F: Enterogastronereactive Escherichia coli pathotype: a genetically heterogeneous emerging foodborne enteropathogen. *FEMS Immunol Med Microbiol* 2016, 82(3):281–298.

6. Frank C, Weber D, Cremer JP, Aukar M, Faben M, Heiden M, Bernard H, Fruth A, Prager R, Spode A, Wadd M, Zoufaly A, Jordan S, Stark K, Krause G: Epidemiological profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. *N Engl J Med* 2011, 365(19):1771–1780.

7. Bielaszewska M, Mielmann A, Zhang W, Köck R, Fruth A, Bauwens A, Peters G, Karch H: Characterisation of the enterohaemorrhagic Escherichia coli associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis* 2011, 11(8):e71–e76.

8. Schuchtez F, Nielsen EM, Froimodt-Müller J, Boisen N, Morabito S, Tozzoli R, Nataro JP, Caprioli A: Characteristics of the enterogastronereactive Shiga toxin/verotoxin-producing Escherichia coli O104:H4 strain causing the outbreak of haemolytic uraemic syndrome in Germany, May to June 2011, *Euro Surveill* 2011, 16(24). Available online: http://www.eursurveillance.org/ViewArticle.aspx?ArticleId=18989.

9. Brauskiewicz E, Thümmer A, Schüdels J, Leimbach A, Liesegang H, Meyer FY, Boelter J, Petersen H, Gottschalk G, Daniel R: Genome sequence analyses of two isolates from the recent Escherichia coli outbreak in Germany reveal the emergence of a new pathotype: Enterogaregative-Haemorrhagic Escherichia coli (EAEHC). *Arch Microbiol* 2011, 193(12):883–891.

10. Morabito S, Karch H, Nartani-Kurkdjian P, Schmidt H, Minelli F, Bingen E, Iyoda S, Tamura K, Itoh K, Izumiya H, Ueno N, Nagata K, Togo M, Terajima J, Prager R, Spode A, Wadl M, Zoufaly A, Jordan S, Stark K, Krause G, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Karch H: Molecular cloning and analysis of genes for production of K5, K7, K12, and K92 capsular polysaccharides in Escherichia coli. *J Bacteriol* 1986, 168(3):1228–1233.

11. Sambrook J, Green RD: Molecular Cloning: A laboratory manual. 3rd edition. New York: Cold Spring Harbor Laboratory Press; 2001.

12. Darling AE, Mau B, Perna NT: ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS One* 2010, 5(11):e111147.

13. Galaxy/CRIS4. In [http://orion.ocris4.de/](http://orion.ocris4.de/).

14. GC Calculato. In [www.genomicsmcplace.com/gc.calc.html](http://www.genomicsmcplace.com/gc.calc.html).

15. Lowe TM, Eddy SR: tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997, 25(12):956–964.

16. Autschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990, 215(3):403–410.

17. Darzentas N: Circoletto: visualizing sequence similarity with Circos. *Bioinformatics* 2010, 26(20):2860–2861.

18. Circoletto. In [http://tools.bat.infspire.org/circoletto/](http://tools.bat.infspire.org/circoletto/).

19. Beutin L, Hammel JA, Strauch E, Reetz J, Dieckmann R, Kelner-Burgos Y, Martin A, Miks, Stockbine NA, Lindstedt BA, Horn D, Monse H, Huetbelt B, Müller L, Stüber K, Reinhardt R: Spread of a distinct Stx2-encoding phage prototype among Escherichia coli O104:H4 strains from outbreaks in Germany, Norway, and Canada, *J Virol* 2012, 86(19):10444–10455.

20. Schirmer B, Nygard K, Erkens HA, Lassen J, Lindstedt BA, Brandal LT, Kopperud G, Aavitsland P: Outbreak of haemolytic uraemic syndrome in Norway caused by stx2-positive Escherichia coli O104:H4 traced to cured mutton sausages. *BMJ Infect Dis* 2008, 8:41.

21. L’Abée-Lund TM, Jørgensen HJ, O’Sullivan K, Bohlin J, Ljungblad G, Granum PE, Lindbäck T: The highly virulent 2006 Norwegian EHEC O103:H25 outbreak strain is related to the 2011 German O104:H4 outbreak strain. *PloS One* 2012, 7(3):e31413.

22. Ahmed SA, Awoiska J, Baldwin C, Bishop-Lilly KA, Biswas B, Broomall S, Chain PS, Chertkov O, Chokoshvili O, Coyne S, Davenport K, Detter JC, Darling AE, Mau B, Perna NT: Molecular Cloning: A laboratory manual. 3rd edition. New York: Cold Spring Harbor Laboratory Press; 2001.

23. Chain PS, Chertkov O, Chokoshvili O, Coyne S, Davenport K, Detter JC, Darling AE, Mau B, Perna NT: ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS One* 2011, 7(3):e22751.

24. Roberts I, Mountford R, High N, Bitter-Suermann D, Jann K, Timmis K, Samadpour M, Whittam TS, Kaul R, Brittnacher M, Miller SI: Comparative genomics of enterohemorrhagic Escherichia coli O145:H28 demonstrates a common evolutionary lineage with Escherichia coli O157:H7. *BMC Genomics* 2014, 15:17.

25. Ogura Y, Ooka T, Ijuichi A, Toh H, Asadulghani M, Oshima K, Kodama T, Aaberge L, Ritchie JM, Raden M, McKeever M, Freeman T, Hayden H, Haugen E, Gillett W, Feng C, Chang J, Beskhelevnya V, Waldor M, Samadpour M, Whittam TS, Kaul R, Brittnacher M, Miller SI: Analysis of the genome of the Escherichia coli O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. *Infect Immun* 2009, 77(3):1713–1721.

26. Eppingen M, Mammel MK, Leckie JE, Ravel J, Cebula TA: Genomic anatomy of Escherichia coli O157:H7 outbreaks. *Proc Natl Acad Sci U S A* 2011, 108(50):20142–20147.

27. Makino K, Yokoyama K, Kubota Y, Yutsudo CH, Kimura S, Kurokawa K, Ishii K, Hattori M, Tatsuno I, Abe H, Iida T, Yamamoto K, Onishi M, Hayashi T, Yusunaga T, Honda T, Sasaki C, Shinagawa H: Complete nucleotide
sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak. *Genes Genet Syst* 1999, 74(5):227–239.

40. Xiong Y, Wang P, Lan R, Ye C, Wang H, Ren J, Jing H, Wang Y, Zhou Z, Bai X, Cui Z, Luo X, Zhao A, Wang Y, Zhang S, Sun H, Wang L, Xu J. A novel Shiga toxin 2-converting bacteriophage was isolated from the Shiga toxin 2-positive strain E. coli O157:H7 reference strain Sakai. *BMC Genomics* 2012, 13:356.

41. Perna NT, 3rd Plunkett G, Burland V, Bau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gogarten JP, Koonin EV, Nelson CE, Plaza-Ramírez E, Rasko DA, Wang Q, Yutö P, Markowitz VM, Hedgpeth ND, Aravind L, Fleischmann RD, Ketchum KA, Deboy RT, et al. Genome sequence of enterohemorrhagic Escherichia coli O157:H7. *Nature* 2001, 409:529–533.

42. Sato T, Shimizu T, Watarai M, Kobayashi M, Kano S, Hamabata T, Takeda Y, Yamazaki S. Genome analysis of a novel Shiga toxin 1 (Stx1)-converting phage which is closely related to Stx2-converting phages but not to other Stx1-converting phages. *J Bacteriol* 2003, 185(13):4966–4971.

43. Sato T, Shimizu T, Watarai M, Kobayashi M, Kano S, Hamabata T, Takeda Y, Yamazaki S. Distinctiveness of the genomic sequence of Shiga toxin 2-converting phage isolated from Escherichia coli O157:H7 Okayama strain as compared to other Shiga toxin 2-converting phages. *Gene* 2003, 309(1):35–48.

44. Su L, Lu CP, Wang Y, Cao DM, Sun JH, Yan YK. Lyssogenic infection of a Shiga toxin 2-converting bacteriophage changes host gene expression, enhances host acid resistance and motility. *J Virol* (Mosk) 2010, 44(1):60–73.

45. Thorvaldóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013, 14(2):178–192.

46. Smith DL, Rooks DJ, Fogg PC, Darby AC, Thomson NR, McCarthy AJ, Allison HE. Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics* 2012, 13:311.

47. Johansen BK, Wasteson Y, Granum PE, Brynestad T. Mosaic structure of Shiga-toxin-2-encoding phages isolated from Escherichia coli O157:H7 indicates frequent gene exchange between lambdoid phage genomes. *Microbiology* 2001, 147(Pt 7):1929–1936.

48. Steyert SR, Sahl JW, Fraser CM, Teel LD, Scheutz F, Rasko DA. Comparative genomics and stx phage characterization of LEE-negative Shiga toxin-producing Escherichia coli. *Front Cell Infect Microbiol* 2012, 2:133.

49. Allison HE. Stx-pherages: drivers and mediators of the evolution of STEC and STEC-like pathogens. *Future Microbiol* 2007, 2(2):165–174.

50. Muniesa M, Blanco JE, De Simón M, Serrano-Moreno R, Blanch AR, Jofre J. Diversity of stx2 converting bacteriophages induced from Shiga-toxin-producing Escherichia coli strains isolated from cattle. *Microbiology* 2004, 150(9):2959–2971.

51. García-Aljaro C, Muniesa M, Jofre J, Blanch AR. Genotypic and phenotypic diversity among induced, stx2-carrying bacteriophages from environmental Escherichia coli strains. *Appl Environ Microbiol* 2009, 75(2):329–336.

52. Werts C, Michel V, Hofnung M, Charbit A. Adsorption of bacteriophage lambda on the LamB protein of Escherichia coli K-12: point mutations in gene J of lambda responsible for extended host range. *J Bacteriol* 1994, 176(4):941–947.

53. Blasche S, Wuchty S, Rajagopala SV, Uetz P. The protein interaction network of bacteriophage lambda with its host, Escherichia coli. *J Virol* 2013, 87(3):12745–12755.

54. Kyle J, Cummings CA, Parker CT, Quiñones B, Vatta P, Newton E, Huynh S, Swinney M, Degoricija L, Barker M, Fontanoz S, Nguyen K, Patel R, Fang R, Tebb R, Penasukiené O, Furtado M, Mandrell RE. Escherichia coli serotype O55:H7 diversity supports parallel acquisition of bacteriophages at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. *J Bacteriol* 2012, 194(8):1885–1896.

55. Imamovic L, Ballesté E, Jofre J, Muniesa M. Quantification of Shiga toxin-converting bacteriophages in wastewater and in fecal samples by real-time quantitative PCR. *Appl Environ Microbiol* 2010, 76(17):5693–5701.

56. Homitzky MA, Mercieca K, Bettelheim KA, Djordjevic SP. Bovine feces from animals with gastrointestinal infections are a source of serologically diverse atypical enteropathogenic Escherichia coli and Shiga toxin-producing E. coli strains that commonly possess intimin. *Appl Environ. Microbiol* 2005, 71(3):940–941. doi:10.1128/AEM.71.3.940-941.2005.

57. Boisen N, Nataro JP and Krogfelt KA. Enterogauggative Escherichia coli and Disease. In Pathogenic Escherichia coli, Molecular and Cellular Microbiology. Edited by Morabito S. Caister Academic Press; 2014:217–230.