Communication

Hybrid Atypical Enteropathogenic and Extraintestinal *Escherichia coli* (aEPEC/ExPEC) BA1250 Strain: A Draft Genome

Danielle D. Munhoz [1,*], Fernanda F. Santos [2,†], Thais Mitsunari [1], Paulo A. Schüroff [1], Waldir P. Elias [1], Eneas Carvalho [1] and Roxane M. F. Piazza [1,*]

[1] Laboratório de Bacteriologia, Instituto Butantan, São Paulo 05503-900, Brazil; thais.mitsunari@butantan.gov.br (T.M.); paulo.schuroff@butantan.gov.br (P.A.S.); waldir.elias@butantan.gov.br (W.P.E.); eneas.carvalho@butantan.gov.br (E.C.)

[2] Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo 04023-062, Brazil; ff.santos@unifesp.br

* Correspondence: dani.dmunhoz@hotmail.com (D.D.M.); roxane@butantan.gov.br (R.M.F.P.); Tel.: +55-112-627-9724 (D.D.M. & R.M.F.P.)

† Present address: Laboratório Alerta, Disciplina de Infectologia, Departamento de Medicina, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo 04039-032, Brazil.

Abstract: Diarrheagenic *Escherichia coli* is the major bacterial etiological agent of severe diarrhea and a major concern of public health. These pathogens have acquired genetic characteristics from other pathotypes, leading to unusual and singular genetic combinations, known as hybrid strains and may be more virulent due to a set of virulence factors from more than one pathotype. One of the possible combinations is with extraintestinal *E. coli* (ExPEC), a leading cause of urinary tract infection, often lethal after entering the bloodstream and atypical enteropathogenic *E. coli* (aEPEC), responsible for death of thousands of people every year, mainly children under five years old. Here we report the draft genome of a strain originally classified as aEPEC (BA1250) isolated from feces of a child with acute diarrhea. Phylogenetic analysis indicates that BA1250 genome content is genetically closer to *E. coli* strains that cause extraintestinal infections, other than intestinal infections. A deeper analysis showed that in fact this is a hybrid strain, due to the presence of a set of genes typically characteristic of ExPEC. These genomic findings expand our knowledge about aEPEC heterogeneity allowing further studies concerning *E. coli* pathogenicity and may be a source for future comparative studies, virulence characteristics, and evolutionary biology.

Keywords: *E. coli*; diarrhea; aEPEC; genome evolution; extraintestinal pathogen

1. Introduction

Diarrhea is the second leading cause of death in children under five years of age, leading approximately 525,000 children to death annually. In low-income countries, children under three years old experience an average of three episodes of serious diarrhea each year, resulting in severe cases of malnutrition and death [1]. Infections of the intestinal tract usually have diarrhea as one of the symptoms and can be caused by a wide variety of bacteria, viruses, and parasites [2], and the main bacterial agent are diarrheagenic *Escherichia coli* (DEC) strains. The frequency of infections caused by these microorganisms can be even higher since in 34% of diarrhea cases the causative agent is not identified [3].

Diarrheagenic *E. coli* evolved by acquiring specific set of characteristics by horizontal gene transfer. The current DEC pathotypes are classified according to the set of acquired genes, and differ in terms of virulence factors, mechanisms of pathogenicity and clinical symptoms [4–6]. Despite the presence of specific virulence factors in each pathogroup, *E. coli* has great genomic plasticity, which has led to the identification of isolates with a combination of virulence characteristics from different pathotypes. These isolates are considered potentially more virulent and are called hybrid pathogenic strains [4,7], as the strain responsible for a large diarrhea outbreak in Germany in 2011 [8–10].

Enteropathogenic *E. coli* (EPEC) is highly prevalent in community settings such as schools and hospitals and is a major cause of childhood diarrhea in developing country. It is estimated that EPEC is responsible for 5–10% of all cases of pediatric diarrhea in countries like Brazil, Chile, Peru, and Iran [11] and, in addition to colonizing humans, this pathogen is also capable of infecting animals such as cattle, dogs, cats, wild boar, and rabbits [12–15]. EPEC is subdivided into typical EPEC (tEPEC) and atypical EPEC (aEPEC), based on the presence of the EPEC adherence factor (EAF) plasmid in the former group and its absence in the latter group [16]. In addition, aEPEC neither express nor produce the bundle-forming pilus (BFP), a type four fimbriae [17–19].

Atypical EPEC strains have been described as an important agent of diarrhea since the 1990s and are currently frequently detected in underdeveloped countries [3,20]. Unlike tEPEC, aEPEC has been detected in patients of all ages and also in immunosuppressed individuals [21–24]. This is an extremely heterogeneous group in terms of virulence factors [25,26], genetic characteristics, in vitro adhesion phenotype, O:H serotypes [27,28], and host type and immunological condition [12,25]. Previous genome assemblies showed that the genome of some aEPEC strains contains different DEC and extraintestinal pathogenic *E. coli* (ExPEC) genetic characteristics [29–31], indicating that this pathotype genome had undergone unique evolutionary events.

In this work, we report the sequencing, assembly, and annotation of the genome of a strain previously classified as aEPEC, BA1250 [17], according to this pathotype genetic markers, as the presence of genes that comprise the pathogenicity island called Locus of enterocyte effacement (LEE) [32]. Paired-end sequencing and trimming of the adapter and small or low-quality sequences were performed. The assembly and annotation of the genome were carried out using the SPAdes and Prokka pipelines, respectively. Interestingly, phylogenetic analysis showed a more recent relationship among BA1250 with *E. coli* strains responsible for urinary infection than with other aEPEC strains. In addition, this strain genome presents a set of genes characteristic of ExPEC, which allows its classification as a hybrid strain.

The genome of BA1250 aEPEC/ExPEC strain will be useful for comparative studies as a genomic reference to allow analyses of singular virulence factors and for studies of molecular and evolutionary biology.

2. Results and Discussion

2.1. Overview of BA1250 Draft Genome Assembly

Approximately 3.6 million fragments of genomic DNA (reads) were sequenced, totalizing approximately 915 million sequenced bases. From this total, adapters sequence, low-quality bases, small reads, and unmatched reads were removed, resulting in approximately 3.4 million reads and 400 million bases, a reduction of approximately 6.9% and 56.9% in the total sequenced, respectively. This value represents the coverage of approximately 77 times, considering an average *E. coli* genome size of 5.1 million bases. The sequencing performed generated data that allow the assembly of a representative genome, as coverage of 30 times is already considered sufficient [33].

The assembled genome has 4,964,536 bases and assembly resulted in 151 contigs, with an N50 of 129,195 bases, and 4717 genes identified after annotation. There were 1261 genes that overlap another gene on the same strand. Comparing to the genome of the prototype tEPEC strain E2348/69 (GCA_000026545.1), there were 83.14% of aligned regions and 79.14% of the genes of the reference strain were found in BA1250. These results indicate that, although it has its particularities, the assembled genome is similar to a reference genome, which suggests that the sequencing and the assembly step could generate material with a good representation of the genomic content of the BA1250 strain. The overall BA1250 draft genome feature (including chromosome and plasmid content) is shown in Table 1.

Table 1. Summary statics of BA1250 draft genome assemblies.

Feature	Value
Contigs	151
GC Content (%)	50.56
Contig L50	9
Contig N50	129,195
Genome length (bp)	4,964,536
Protein coding sequences (CDS)	4969
Annotated genes	4717
Transfer RNA (tRNA)	87
Ribosomal RNA (rRNA)	12

2.2. Plasmid Content

The results obtained by Illumina sequencing (Illumina, Inc., San Diego, CA, USA) and the SCAPP [34] assembler allowed the identification of a complete plasmid sequence. The plasmid, assembled to a unique contig, has 154,710 base pairs, with 158 CDS and 158 genes identified after annotation carried out using the Prokka pipeline.

Plasmid classification into Incompatibility (Inc) groups is based on the sequence of the replication initiation (Rep) protein. Using Plasmid Finder software [35], we have identified, with 96,5% of identity, the Inc group IncFIB/IncFII in the assembled plasmid. Plasmids classified as IncFIB/IncFII are found in bacteria from different animals and human hosts, and the strains belong to several phylogroup, as A, B1, B2, D, and E [36]. The F plasmid is a primary example of a large group of conjugative plasmids, widespread in *E. coli.* It is worth mentioning that virulence-associated traits of *E. coli* are almost exclusively found on IncF-family plasmids, making these plasmids of great clinical relevance [37,38]. The majority of proteins comprised in F-family plasmid-mediated bacterial conjugation are expressed from a polycistronic *tra* operon in the plasmid transfer (*tra*) region that contains basically the same organization of conjugation-related genes [39,40]. In the BA1250 plasmid sequence, based on an additional analysis carried out with Abricate [41] were identified nine *tra* genes, as *traA, traC, traD, traJ, traM, traQ, traS, traV,* and *traY*.

Among the 158 genes present in this plasmid, is the *vat* gene, a highly prevalent and tightly regulated immunogenic serine protease autotransporter protein of Enterobacteri-aceae (SPATE) usually secreted by ExPEC during infection and considered as an ExPEC marker [42–44]. It is also present in the plasmid sequence, seven genes of the *fae* operon, which encodes subunits of chaperone-usher assembled family K88 fimbriae [45].

2.3. Phylogenetic Analyses

To understand the general genomic relationship of BA1250 with other *E. coli* isolates, phylogenetic analyses were performed based on the genome of 13 aEPEC strains, 4 DEC prototypes, and 16 ExPEC strains. Phylogeny tree indicated that BA1250 shares a more recent common ancestor with *E. coli* strains isolated from extraintestinal infections than with other aEPEC strains (Figure 1A). In addition to the completely different site of isolation, this analysis demonstrates that BA1250 and ExPEC share genetic characteristics and are more closely related to each other than BA1250 with the other strains of aEPEC selected for the analysis.
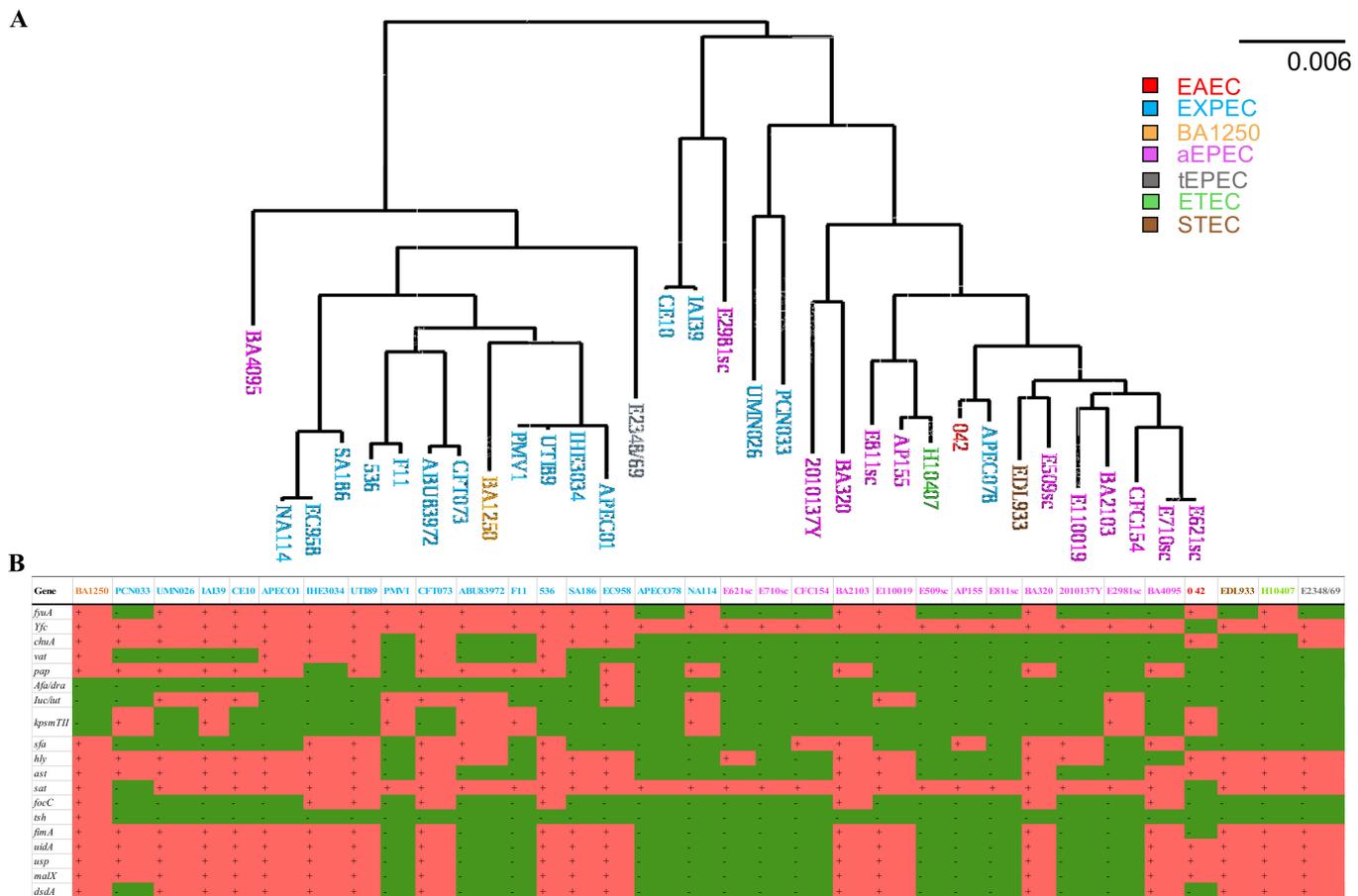
A

B

Figure 1. BA1250 phylogenetic analysis. (**A**) The phylogenetic tree was built with 13 atypical enteropathogenic *E. coli* (aEPEC) strains, 4 diarrheagenic *Escherichia coli* (DEC) prototypes, and 16 extraintestinal pathogenic *E. coli* (ExPEC) strains genome. Strains are colored according to pathogenic *E. coli* group, as blue for ExPEC strains, brown for Shiga-toxin-producing *E. coli* (STEC) prototype strain EDL933, pink for aEPEC strains, red for enteroaggregative *E. coli* (EAEC) prototype strain 042, gray for typical enteropathogenic *E. coli* (tEPEC) prototype strain E2348/69, green for enterotoxigenic *E. coli* (ETEC) prototype strain H10407 and orange for the aEPEC strain BA1250. (**B**) Specific ExPEC gene survey among selected ExPEC, aEPEC and prototypes DEC strains. All strains names are colored as described for phylogenetic tree. Red squares represent presence of gene and Green squares represent its absence.

Hybrid *E. coli* pathotypes are being classified as emerging public health threats with enhanced virulence due to genetic features from different pathotypes. To determine whether BA1250 could be classified as a hybrid strain, a deeper genetic survey was performed. Most of DEC pathotypes possess a set of characteristic genes that are targeted in pathotype classification, epidemiologic studies as well as in diagnostic tools. Although their isolation site designates ExPEC strains, some groups of genes are often employed as ExPEC markers, since it does not have exclusive and well-defined core of characteristic genes. Research groups use distinct set of genes, usually four or five genes, to study virulence factor among ExPEC strains [29,42–44]. In this study we have selected 19 genes of the most common virulence features of this pathotype to investigate whether BA1250 may be classified as a hybrid strain. A survey in BA1250 genome revealed that, in fact, this is a hybrid strain, since it possesses 16 genes that are ExPEC markers (Figure 1B).

Considering that BA1250 has the genetic markers of ExPEC and may now be classified as hybrid strain, we applied the Average Nucleotide Identity (ANI) [46] and compared its genome with the 16 ExPEC, 12 aEPEC and 4 DEC prototypes strains used in the phylogenetic tree. As presumed, the comparative assay revealed a high similarity between each strain from a specific pathotype, as aEPEC and ExPEC, as depicted by the hierarchical clustering based on ANI values (Figure 2A). Likewise, there were reduced similarities

between the strains from different pathotypes. In accordance with the phylogenetic analysis, BA1250 has more similarity with the ExPEC group, showing a higher level of identity with the extraintestinal pathogen compared to the intestinal pathogen (Figure 2B).
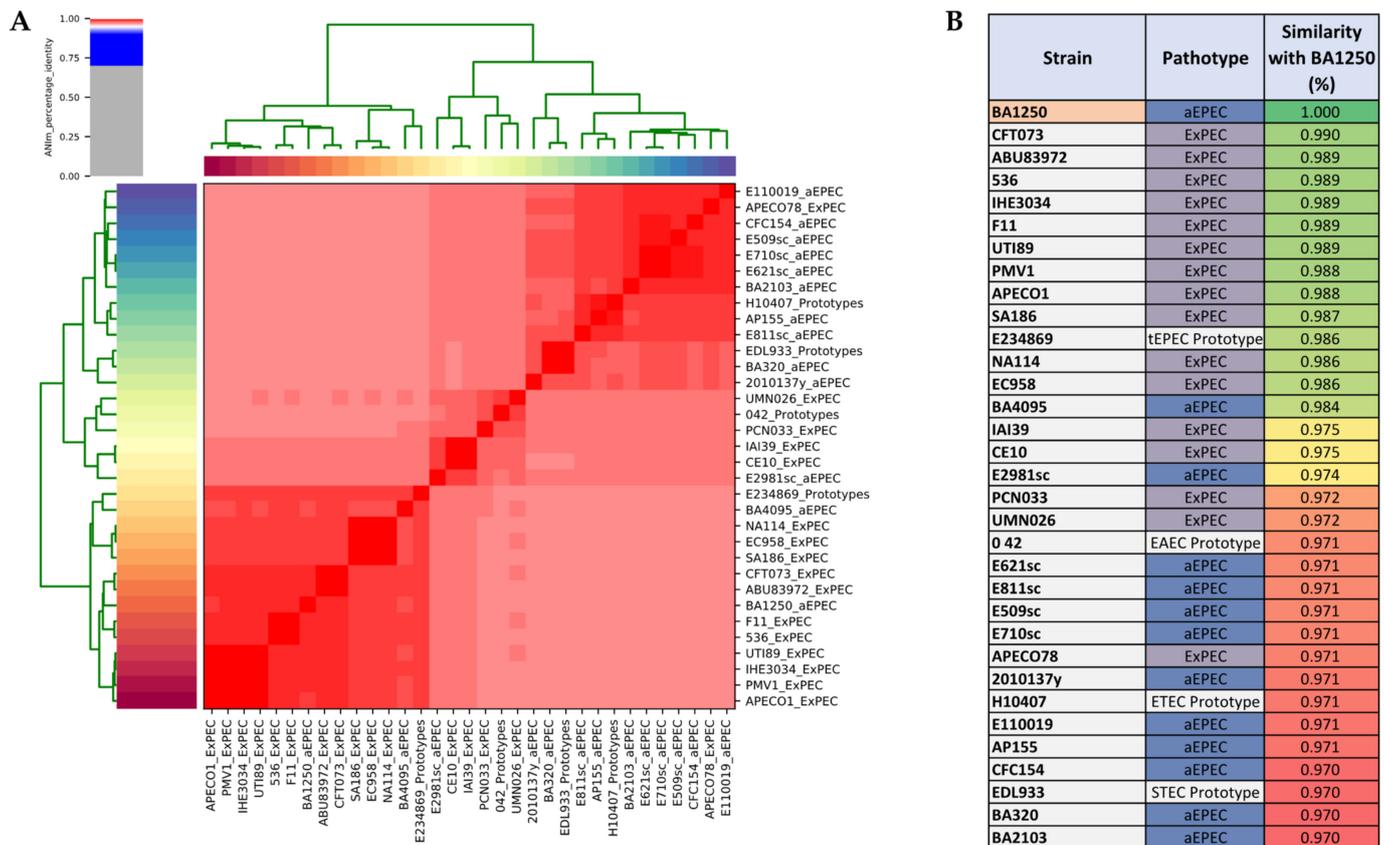

Strain	Pathotype	Similarity with BA1250 (%)
BA1250	aEPEC	1.000
CFT073	ExPEC	0.990
ABU83972	ExPEC	0.989
536	ExPEC	0.989
IHE3034	ExPEC	0.989
F11	ExPEC	0.989
UTI89	ExPEC	0.989
PMV1	ExPEC	0.988
APECO1	ExPEC	0.988
SA186	ExPEC	0.987
E234869	tEPEC Prototype	0.986
NA114	ExPEC	0.986
EC958	ExPEC	0.986
BA4095	aEPEC	0.984
IAI39	ExPEC	0.975
CE10	ExPEC	0.975
E2981sc	aEPEC	0.974
PCN033	ExPEC	0.972
UMN026	ExPEC	0.972
0 42	EAEC Prototype	0.971
E621sc	aEPEC	0.971
E811sc	aEPEC	0.971
E509sc	aEPEC	0.971
E710sc	aEPEC	0.971
APECO78	ExPEC	0.971
2010137y	aEPEC	0.971
H10407	ETEC Prototype	0.971
E110019	aEPEC	0.971
AP155	aEPEC	0.971
CFC154	aEPEC	0.970
EDL933	STEC Prototype	0.970
BA320	aEPEC	0.970
BA2103	aEPEC	0.970

Figure 2. Comparison of Average Nucleotide Identity between each of the 33 strains used in this study. (**A**). The cells in the heatmap are colored according to an Average Nucleotide Identity (ANI) value demonstrated at the top left box. The dendrograms (above and on the left side), correspond to the results of the clustering of the ANI values between the used strains. (**B**). The figure demonstrates the percentage of similarity specifically with BA1250, colored according to a colour scale from green to red.

Unlike BA1250, the most notorious hybrid strain was originated from two DEC pathotypes, instead of DEC and ExPEC. This Shiga-toxin-producing *E. coli* (STEC)/enteroaggregative *E. coli* (EAEC) strain was responsible for a large outbreak with numerous hemolytic-uremic syndrome (HUS) cases and deaths in Germany in 2011 [10]. The strain was initially classified as STEC due to the presence of *stx* genes encoding Shiga toxin, but were notably more virulent, since it led to a higher frequency of HUS cases and even deaths. Genome sequencing and phylogeny analysis revealed a significant genetic similarity with EAEC strains and the presence of classical EAEC virulence genes as *aggA* and *aggR*.

Some DEC/ExPEC hybrid strains also have been identified, for instance the STEC/ExPEC O80:H2 hybrid reported to cause HUS and bacteremia [47]. This specific strain was successively isolated from the patient stool and blood sample, showing its extreme and diverse virulence capacity. Genetic analysis revealed the presence of the typical STEC *stx* genes and some ExPEC genes like *iucC* and *hlyF* [48]. Similarly, a significant number of strains were isolated from hospitalized patients and further characterized according the genetic and phenotypic characteristics in in vitro assays as DEC/ExPEC hybrids [49].

Concerning another EPEC/ExPEC hybrid, isolated from a case of a patient with severe protracted diarrhea, bacteremia and multiorgan dysfunction, genome sequencing revealed that this infection was caused by an ExPEC strain that also featured distant orthologs of genes characteristic of EPEC [50]. This strain displayed some characteristic of ExPEC,

including complete sets of genes for P fimbriae, hemolysin, and the *kpsM* gene used to identify group II capsule loci [51]. The study also revealed that this isolate has the genetic markers for EPEC, including the presence of LEE genes and the absence of *stx1* and *stx2* genes encoding Shiga toxins, which are found in addition to the LEE in enterohemorrhagic *E. coli* (EHEC) strains [16]. However, some essential LEE genes were not found or displayed low similarity, raising doubt as to whether the characteristic EPEC Type 3 secretion system (T3SS) is fully functional in this isolate.

2.4. BA1250: A Hybrid Strain

The ability of a specific strain to cause disease relay in a specific combination of genes encoding virulence factors. The core genome shared by *E. coli* strains corresponds to approximately 40% of its average 5000 genes, while the total number of genes that exist in all *E. coli* strains is predicted to exceed 15,000 [52,53]. Each diarrheal and extraintestinal *E. coli* pathotype possess a distinguished genomic feature that illustrates the phenotypical characteristics.

In this study we described a number of genomic features from BA1250, a strain that were initially classified as aEPEC. This classification was based on the EPEC unique characteristic that is the presence of a pathogenicity island named as Locus of Enterocyte Effacement (LEE) [32]. This pathogenicity island is composed of 41 genes and some of them are used as representative gene markers for EPEC, such as *eae*, encoding the adhesin intimin, and *tir*, encoding the intimin translocated receptor. This pathogenicity island is likely to encode almost all of the genes necessary to produce an intestinal attaching/effacing (A/E) lesion, the histopathological lesion characteristic of EPEC infection. In addition to *eae* and *tir*, LEE comprises genes that include a type III secretion system (TTSS), several secreted proteins (*espB*, *espD*, and *espF*), and their chaperones [54–56]. Distinct from EPEC, a strain is usually classified as ExPEC according to its extraintestinal site of isolation. While several virulence factors are associated with ExPEC, until now there is no core set of genes that can be used to definitively differentiate these pathotypes. However, over the last years, authors have observed some group of genes more frequently identified among ExPEC strains, thus being applied as genetic markers [29,42–44,57]. These set of genes vary among four or five genes, such as *fyuA*, *yfc*, *chuA*, and *vat*, which, among others, were identified in the BA1250 genome, allowing its classification as a hybrid strain.

E. coli is a highly diverse group, yet still closely related microorganism. Until now, there is not a consensus regarding the exact evolution process of DEC strains [58,59]; thus, it is still not feasible to classify which part of genome belongs to each pathotype. Nevertheless, as we mentioned, some groups of genes are described for each pathotype and applied as genetic marker, as LEE genes are for EPEC and 19 genes of ExPEC selected for this study. In Figure 3 we highlighted some of the LEE genes and ExPEC genome block in BA1250 genome. All LEE genes are grouped as a pathogenic island present in the EPEC strains chromosome (revised in 25). Except from *vat* gene that were found in the BA1250 plasmid, all other 15 ExPEC genes, were also identified in the chromosome, where genes *fimA*, *focC*, *sfaH*, *hlyE*, *chuA*, *tsh*, *malX*, and *papC* are part of different pathogenicity islands according to literature and predicted by Island Viewer 4 [60–62].

Considering the emergence and public health significance of hybrid pathotypes, this new pathogenic group should be studied in closer details to better understand the hybrid infection. The genome of BA1250 is a very interesting model of a hybrid strain and this sequencing investigation allows future comparison analyzes.

Figure 3. Illustrative BA1250 draft genome (including chromosome and plasmid contents). Representative LEE genes from EPEC are stained in pink. The selected ExPEC genes for the study are stained in blue. * genes present in chromosome, † gene present in plasmid and # genes that are part of pathogenicity islands. Contigs are represented as gray arrows in the inner circle and are classified with respect to their size. Colored lines represent genes in the positive and negative-sense DNA strand into the outer and inner circle, respectively.

3. Materials and Methods

3.1. Strain Isolation and Identification

BA1250 strain analyzed in this study was isolated from feces in a case-control epidemiologic survey from children with diarrhea in Salvador, Brazil. The strain was classified as atypical enteropathogenic *E. coli* (aEPEC, *eae*+/EAF−/*stx*−/BFP−), displaying the localized adherence-like (LAL) pattern on HEp-2 cells [17,63]. The strain was stored at −80 °C in Luria Bertani broth (LB) with 25% glycerol and was routinely grown in LB for 18 h at 37 °C.

3.2. DNA Extraction, Sequencing Library, and Read Filtering

For BA1250 genome sequencing, the strain was grown in LB broth for 18 h at 37 °C, and DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega, Madison, WI, USA), according to the manufacturer's protocol. The sequencing was performed using Illumina technology on the Illumina Hiseq1500 platform, with kits from the same company and following the manufacture's guidelines. The genomic library was prepared using the HiSeq Rapid SBS Kit v2 library kit, and paired-end sequencing was performed with 2 × 250 sequencing cycles. The resulted sequences were treated for the removal of adapter, low-quality bases, and small reads, using the Trimmomatic tool [64].

3.3. Genome Assembly and Annotation

Paired-end reads were assembled using SPAdes version 3.11.1 [65] and QUAST tool was applied to analyze the assembled genome content and align it to a reference

genome E2348/69 (GCA_000026545.1) [66]. Finally, the assembled genome was subjected to automated annotation by Prokka program version 1,12 [67]. The whole-genome sequences (WGS) were deposited in the GenBank database under the accession number JADPBX000000000, BioProject and SRA data PRJNA678986.

3.4. Plasmid Identification

Reads were assembled using SCAPP [34] and PlamisFinder tool [35] was applied to analyze the assembled genome content. The assembled plasmid was subjected to automated annotation by Prokka program version 1,12 [67] that enabled protein identification.

3.5. Phylogenetic Analysis

The assembled genome from BA1250 was submitted to the comprehensive genome analysis service at Pathosystems Resource Integration Center (PATRIC) [68], where a phylogenetic analysis was carried out, on the "Phylogenetic Tree Building" using default option "Codon Tree". Phylogeny tree were edit with TreeDyn 198.3 at Methods et algorithmes pour la bio-informatique from LIRMM [69–71]. The analysis was performed based on the genome of 13 aEPEC strains (BA1250; Strain 2010137Y; AP155 (HE-8); E110019; BA4095; BA2103; BA320; E710sc; E621sc; E811sc; E2981sc; E509sc; CFC_154), 4 DEC prototypes (ETEC strain H10407; EHEC strain EDL933; EAEC strain 042; tEPEC strain E2348/69) and 16 ExPEC strains (APEC078; PCN033; UMN026; IAI39; CE10; APEC01; IHE3034; UTI89; PMV1; CFT073; ABU83972; F11; 536; SA186; EC958; NA114). All aEPEC strains were isolated from human samples, except strain AP155 (HE-8) that was isolated from dog and strain CFC_154 that was isolated from cow. Gene specific analyses were carried out with Sublime Text editor.

The different genome comparison was carried out with Average Nucleotide Identity (ANI), using the software pyany [46]. This is a category of computational analysis that usually involves the fragmentation of genome sequences, followed by nucleotide sequence search, alignment, and identity calculation.

3.6. Pathogenic Island Predictiom

The assembled genome from BA1250 was submitted to the server of Island Viewer [63]. IslandViewer is a computational tool that integrates four different genomic island prediction methods: IslandPick, IslandPath-DIMOB, SIGI-HMM, and Islander.

4. Conclusions

This study clearly demonstrated that the genetic background of BA1250 closely resembled to ExPEC strains, which could indicate that this is a hybrid strain and may be of greater virulence compared to other aEPEC, since it possesses machinery to possibly infect different niches other than the intestine. This comparative analysis provided a comprehensive understanding of an aEPEC strain genome and closely related species that allow future studies on hybrid strains.

Author Contributions: Conceptualization, D.D.M. and R.M.F.P.; methodology, D.D.M., F.F.S., T.M., P.A.S. and E.C.; software, E.C.; validation, D.D.M., F.F.S. and E.C.; formal analysis, D.D.M. and R.M.F.P.; resources, R.M.F.P. and W.P.E.; data curation, D.D.M.; writing—original draft preparation, D.D.M.; writing—review and editing, F.F.S., E.C., W.P.E. and R.M.F.P.; supervision, R.M.F.P.; project administration, W.P.E. and R.M.F.P.; funding acquisition, W.P.E. and R.M.F.P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data underlying this article are available in the GenBank database under the accession number JADPBX000000000, BioProject and SRA data PRJNA678986, at https://www.ncbi.nlm.nih.gov/genbank/ (accessed on 15 January 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization (WHO). Diarrhoeal Disease. 2017. Available online: https://www.who.int/en/news-room/fact-sheets/detail/diarrhoeal-disease (accessed on 10 February 2021).
2. Ashworth, A. Treatment of severe malnutrition. *J. Pediatr. Gastroenterol. Nutr.* **2001**, *32*, 516–518. [CrossRef]
3. Lanata, C.F.; Fischer-Walker, C.L.; Olascoaga, A.C.; Torres, C.X.; Aryee, M.J.; Black, R.E. Global causes of diarrheal disease mortality in children <5 years of age: A systematic review. *PLoS ONE* **2013**, *8*, e72788. [CrossRef] [PubMed]
4. Croxen, M.A.; Law, R.J.; Scholz, R.; Keeney, K.M.; Wlodarska, M.; Finlayet, B.B. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* **2013**, *26*, 822–880. [CrossRef] [PubMed]
5. Kaper, J.B.; Nataro, J.P.; Mobley, H.L.T. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2004**, *2*, 123–140. [CrossRef] [PubMed]
6. Nataro, J.P.; Kaper, J.B. Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* **1998**, *11*, 142–201. [CrossRef]
7. Santos, A.C.M.; Santos, F.F.; Silva, R.M.; Gomes, T.A.T. Diversity of hybrid- and hetero-pathogenic *Escherichia coli* and their potential implication in more severe diseases. *Front. Cell Infect. Microbiol.* **2020**, *10*, 339–350. [CrossRef]
8. Brzuszkiewicz, E.; Thürmer, A.; Schuldes, J.; Leimbach, A.; Liesegang, H.; Meyer, F.; 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: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Arch. Microbiol.* **2011**, *193*, 883–891. [CrossRef]
9. Mellmann, A.H.D.; Harmsen, D.; Cummings, C.A.; Zentz, E.B.; Leopold, S.R.; Rico, A.; Prior, K.; Szczepanowski, R.; Ji, Y.; Zhang, W.; et al. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS ONE* **2011**, *6*, e22751. [CrossRef]
10. Rasko, D.A.; Webster, D.R.; Sahl, J.W.; Bashir, A.; Boisen, N.; Scheutz, F.; Paxinos, E.E.; Sebra, R.; Chin, C.; Iliopoulos, D.; et al. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N. Engl. J. Med.* **2011**, *365*, 709–717. [CrossRef]
11. Ochoa, T.J.; Barletta, F.; Contreras, C.; Mercado, E. New insights into the epidemiology of enteropathogenic *Escherichia coli* infection. *Trans. R. Soc. Trop. Med. Hyg.* **2008**, *102*, 852–856. [CrossRef]
12. Hernandes, R.T.; Elias, W.P.; Vieira, M.A.M.; Gomes, T.A.T. An overview of atypical enteropathogenic *Escherichia coli*. *FEMS Microbiol. Lett.* **2009**, *297*, 137–149. [CrossRef]
13. He, T.; Wang, Y.; Qian, M.; Wu, C. Mequindox resistance and in vitro efficacy in animal-derived *Escherichia coli* strains. *Vet. Microbiol.* **2015**, *177*, 341–346. [CrossRef]
14. Singh, P.; Sha, Q.; Lacher, D.W.; del Valle, J.; Mosci, R.E.; Moore, J.A.; Scribner, K.T.; Manning, S.D. Characterization of enteropathogenic and Shiga toxin-producing *Escherichia coli* in cattle and deer in a shared agroecosystem. *Front. Cell Infect. Microbiol.* **2015**, *5*, 29–42. [CrossRef]
15. Bertelloni, F.; Cilia, G.; Bogi, S.; Ebani, V.V.; Turini, L.; Nuvoloni, R.; Cerri, D.; Fratini, F.; Turchi, B. Pathotypes and antimicrobial susceptibility of *Escherichia coli* isolated from wild boar (*Sus scrofa*) in Tuscany. *Animals* **2020**, *10*, 744. [CrossRef]
16. Kaper, J.B. Defining EPEC. *Rev. Microbiol.* **1996**, *27*, 130–133.
17. Abe, C.M.; Trabulsi, L.R.; Blanco, J.; Blanco, M.; Dahbi, G.; Blanco, J.E.; Mora, A.; Franzolin, M.R.; Taddei, C.R.; Martinez, M.B.; et al. Virulence features of atypical enteropathogenic *Escherichia coli* identified by the *eae*+ EAF-negative *stx*-genetic profile. *Diagn. Microbiol. Infect. Dis.* **2009**, *64*, 357–365. [CrossRef]
18. Nara, J.M.; Cianciarullo, A.M.; Culler, H.F.; Bueris, V.; Horton, D.S.P.Q.; Menezes, M.A.; Franzolin, M.R.; Elias, W.P.; Piazza, R.M.F. Differentiation of typical and atypical enteropathogenic *Escherichia coli* using colony immunoblot for detection of bundle-forming pilus expression. *J. Appl. Microbiol.* **2010**, *109*, 35–43. [CrossRef]
19. Gomes, T.A.; Elias, W.P.; Scaletsky, I.C.A.; Guth, B.E.C.; Rodrigues, J.F.; Piazza, R.M.F.; Ferreira, L.C.S.; Martinez, M.B. Diarrheagenic *Escherichia coli*. *Braz. J. Microbiol.* **2016**, *47*, 3–30. [CrossRef]
20. Ochoa, T.J.; Contreras, C.A. Enteropathogenic *Escherichia coli* infection in children. *Curr. Opin. Infect. Dis.* **2011**, *24*, 478–483. [CrossRef]
21. Assis, F.E.A.; Wolf, S.; Surek, M.; De Toni, F.; Souza, E.M.; Pedrosa, F.O.; Farah, S.M.S.S.; Picheth, G.; Fadel-Picheth, C.M.T. Impact of Aeromonas and diarrheagenic *Escherichia coli* screening in patients with diarrhea in Paraná, Southern Brazil. *J. Infect. Dev. Ctries.* **2014**, *8*, 1609–1614. [CrossRef]
22. Dias, R.C.B.; dos Santos, B.C.; dos Santos, L.F.; Vieira, M.A.; Yamatogi, R.S.; Mondelli, A.L.; Sadatsune, T.; Sforcin, J.M.; Gomes, T.A.T.; Hernandes, R.T. Diarrheagenic *Escherichia coli* pathotypes investigation revealed atypical enteropathogenic *E. coli* as putative emerging diarrheal agents in children living in Botucatu, São Paulo State, Brazil. *APMIS* **2016**, *124*, 299–308. [CrossRef]
23. Gomes, T.A.; Irino, K.; Girão, D.M.; Girão, V.B.C.; Guth, B.E.C.; Vaz, T.M.I.; Moreira, F.C.; Chinarelli, S.H.; Vieira, M.A.M. Emerging enteropathogenic *Escherichia coli* strains? *Emerg. Infect. Dis.* **2004**, *10*, 1851–1855. [CrossRef]
24. Lozer, D.M.; Souza, T.B.; Monfardini, M.V.; Vicentini, F.; Kitagawa, S.S.; Scaletsky, I.C.A.; Spano, L.C. Genotypic and phenotypic analysis of diarrheagenic *Escherichia coli* strains isolated from Brazilian children living in low socioeconomic level communities. *BMC Infect. Dis.* **2013**, *13*, 418. [CrossRef]

25. Trabulsi, L.R.; Keller, R.; Gomes, T.A.T. Typical and atypical enteropathogenic *Escherichia coli*. *Emerg. Infect. Dis.* **2002**, *8*, 508–513. [CrossRef]

26. Bando, S.Y.; Andrade, F.B.; Guth, B.E.C.; Elias, W.P.; Moreira-Filho, C.A.; de Castro, A.F.P. Atypical enteropathogenic *Escherichia coli* genomic background allows the acquisition of non-EPEC virulence factors. *FEMS Microbiol. Lett.* **2009**, *299*, 22–30. [CrossRef]

27. Mora, A.; Blanco, M.; Yamamoto, D.; Dahbi, G.; Blanco, J.E.; López, C.; Alonso, M.P.; Vieira, M.A.M.; Hernandes, R.T.; Abe, C.M.; et al. HeLa-cell adherence patterns and actin aggregation of enteropathogenic *Escherichia coli* (EPEC) and Shiga-toxin-producing *E. coli* (STEC) strains carrying different *eae* and *tir* alleles. *Int. Microbiol.* **2009**, *12*, 243–251. [CrossRef]

28. Tozzoli, R.; Scheutz, F. Diarrhoeagenic *Escherichia coli* infections in humans. In *Pathogenic Escherichia coli*; Morabito, S., Ed.; Caister Academic Press: Pooler, UK, 2014; pp. 1–18. ISBN 978-1-908230-37-9.

29. Valiatti, T.B.; Santos, F.F.; Santos, A.C.M.; Nascimento, J.A.S.; Silva, R.M.; Carvalho, E.; Sinigaglia, R.; Gomes, T.A.T. Genetic and virulence characteristics of a hybrid atypical enteropathogenic and uropathogenic *Escherichia coli* (aEPEC/UPEC) Strain. *Front. Cell Infect. Microbiol.* **2020**, *10*, 492. [CrossRef]

30. Hazen, T.H.; Daugherty, S.C.; Shetty, A.C.; Nataro, J.P.; Rasko, D.A. Transcriptional variation of diverse enteropathogenic *Escherichia coli* isolates under virulence-inducing conditions. *mSystems* **2017**, *2*, e00024-17. [CrossRef]

31. Dutta, S.; Pazhani, G.P.; Nataro, J.P.; Ramamurthy, T. Heterogenic virulence in a diarrheagenic *Escherichia coli*: Evidence for an EPEC expressing heat-labile toxin of ETEC. *Int. J. Med. Microbiol.* **2015**, *305*, 47–54. [CrossRef]

32. McDaniel, T.K.; Jarvis, K.G.; Donnenberg, M.S.; Kaper, J.B. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1664–1668. [CrossRef]

33. Kisand, V.; Lettieri, T. Genome sequencing of bacteria: Sequencing, de novo assembly and rapid analysis using open source tools. *BMC Genom.* **2013**, *14*, 211. [CrossRef] [PubMed]

34. Pellow, D.; Zorea, A.; Probst, M.; Furman, O.; Segal, A.; Mizrahi, I.; Shamir, R. SCAPP: An algorithm for improved plasmid assembly in metagenomes. *bioRxiv* **2020**, *1*, 903252. [CrossRef]

35. Carattoli, A.; Zankari, E.; Garcia-Fernandez, A.; Voldby Larsen, M.; Lund, O.; Villa, L.; Aarestrup, F.M.; Hasman, H. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **2014**, *58*, 3895–3903. [CrossRef] [PubMed]

36. Montealegre, M.C.; Talavera Rodríguez, A.; Roy, S.; Hossain, M.I.; Islam, M.A.; Lanza, V.F.; Julian, T.R. High genomic diversity and heterogenous origins of pathogenic and antibiotic-resistant *Escherichia coli* in household settings represent a challenge to reducing transmission in low-income settings. *mSphere* **2020**, *5*, e00704-19. [CrossRef]

37. Johnson, T.J.; Nolan, L.K. Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **2009**, *73*, 750–774, Erratum in: *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 477–478. [CrossRef]

38. Fernandez-Lopez, R.; Toro, M.; Moncalian, M.; Garcillan-Barcia, M.P.; de la Cruz, F. Comparative genomics of the conjugation region of F-like plasmids: Five shades of F. *Front. Mol. Biosci.* **2016**, *71*. [CrossRef]

39. Frost, L.S.; Ippen-Ihler, K.; Skurray, R.A. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **1994**, *58*, 162–210. [CrossRef]

40. Lawley, T.D.; Klimke, W.A.; Gubbins, M.J.; Frost, L.S. F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* **2003**, *224*, 1–15. [CrossRef]

41. Seemann, T. Abricate, Github. Available online: https://github.com/tseemann/abricate (accessed on 23 March 2021).

42. Pitout, J.D. Extraintestinal pathogenic *Escherichia coli*: A combination of virulence with antibiotic resistance. *Front. Microbiol.* **2012**, *9*. [CrossRef]

43. Dale, A.P.; Woodford, N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. *J. Infect.* **2015**, *71*, 615–626. [CrossRef]

44. Nichols, K.B.; Totsika, M.; Moriel, D.G.; Lo, A.W.; Yang, J.; Wurpel, D.J.; Rossiter, A.E.; Strugnell, R.A.; Henderson, I.R.; Ulett, G.C.; et al. Molecular Characterization of the vacuolating autotransporter toxin in uropathogenic *Escherichia coli*. *J. Bacteriol.* **2016**, *198*, 1487–1498. [CrossRef]

45. Werneburg, G.T.; Thanassi, D.G. Pili Assembled by the chaperone/usher pathway in *Escherichia coli* and *Salmonella*. *EcoSal Plus* **2018**, *8*, 10. [CrossRef]

46. Pritchard, L.; Glover, R.H.; Humphris, S.; Elphinstoneb, J.G.; Toth, I.K. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal. Methods* **2016**, *8*, 12–24. [CrossRef]

47. Mariani-Kurkdjian, P.; Lemaître, C.; Bidet, P.; Perez, D.; Boggini, L.; Kwon, T.; Bonacorsi, S. Haemolytic-uraemic syndrome with bacteraemia caused by a new hybrid *Escherichia coli* pathotype. *New Microbes New Infect.* **2014**, *2*, 127–131. [CrossRef]

48. Bielaszewska, M.; Schiller, R.; Lammers, L.; Bauwens, A.; Fruth, A.; Middendorf, B.; Schmidt, M.A.; Tarr, P.I.; Dobrindt, U.; Karch, H.; et al. Heteropathogenic virulence and phylogeny reveal phased pathogenic metamorphosis in Escherichia coli O2:H6. *EMBO Mol. Med.* **2014**, *6*, 347–357. [CrossRef]

49. Toval, F.; Köhler, C.D.; Vogel, U.; Wagenlehner, F.; Mellmann, A.; Fruth, A.; Schmidt, M.A.; Karch, H.; Bielaszewska, M.; Dobrindt, U. Characterization of *Escherichia coli* isolates from hospital inpatients or outpatients with urinary tract infection. *J. Clin. Microbiol.* **2014**, *52*, 407–418. [CrossRef]

50. Kessler, R.; Nisa, S.; Hazen, T.H.; Horneman, A.; Amoroso, A.; Rasko, D.A.; Donnenberg, M.S. Diarrhea, bacteremia and multiorgan dysfunction due to an extraintestinal pathogenic *Escherichia coli* strain with enteropathogenic *E. coli* genes. *Pathog. Dis.* **2015**, *73*, ftv076. [CrossRef]

51. Johnson, J.R.; O'Bryan, T.T. Detection of the *Escherichia coli* group 2 polysaccharide capsule synthesis gene *kpsM* by a rapid and specific PCR-based assay. *J. Clin. Microbiol.* **2004**, *42*, 1773–1776. [CrossRef]
52. Rasko, D.A.; Rosovitz, M.J.; Myers, G.S.; Mongodin, E.F.; Fricke, W.F.; Gajer, P.; Crabtree, J.; Sebaihia, M.; Thomson, N.R.; Chaudhuri, R.; et al. The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* **2008**, *190*, 6881–6893. [CrossRef]
53. Touchon, M.; Hoede, C.; Tenaillon, O.; Barbe, V.; Baeriswyl, S.; Bidet, P.; Bingen, E.; Bonacorsi, S.; Bouchier, C.; Bouvet, O.; et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **2009**, *5*, e1000344. [CrossRef] [PubMed]
54. Knutton, S.; Baldwin, T.; Williams, P.H.; McNeish, A.S. Actin accumulation at sites of bacterial adhesion to tissue culture cells: Basis of a new diagnostic test for enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect. Immun.* **1989**, *57*, 1290–1298. [CrossRef] [PubMed]
55. Donnenberg, M.S.; Yu, J.; Kaper, J.B. A second chromosomal gene necessary for intimate attachment of enteropathogenic *Escherichia coli* to epithelial cells. *J. Bacteriol.* **1993**, *175*, 4670–4680. [CrossRef] [PubMed]
56. Garmendia, J.; Frankel, G.; Crepin, V.F. Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: Translocation, translocation, translocation. *Infect. Immun.* **2005**, *73*, 2573–2585. [CrossRef] [PubMed]
57. Spurbeck, R.R.; Dinh, P.C.; Walk, S.T.; Stapleton, A.E.; Hooton, T.M.; Nolan, L.K.; Kim, K.S.; Johnson, J.R.; Mobley, H.L.T. *Escherichia coli* isolates that carry *vat*, *fyuA*, *chuA*, and *yfcV* efficiently colonize the urinary tract. *Infect. Immun.* **2012**, *80*, 4115–4122. [CrossRef]
58. Castillo, A.; Eguiarte, L.E.; Souza, V. A genomic population genetics analysis of the pathogenic enterocyte effacement island in *Escherichia coli*: The search for the unit of selection. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1542–1547. [CrossRef]
59. Mohammadzadeh, M.; Oloomi, M.; Bouzari, S. Genetic evaluation of locus of enterocyte effacement pathogenicity island (LEE) in enteropathogenic *Escherichia coli* isolates (EPEC). *Iran. J. Microbiol.* **2013**, *5*, 345–349.
60. Parreira, V.R.; Gyles, C.L. A novel pathogenicity island integrated adjacent to the *thrW* tRNA gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. *Infect. Immun.* **2003**, 5087–5096. [CrossRef]
61. Schmidt, H.; Hensel, M. Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.* **2004**, *17*, 14–56, Erratum in: *Clin. Microbiol. Rev.* **2006**, *19*, 257. [CrossRef]
62. Bertelli, C.; Laird, M.R.; Williams, K.P.; SFU Research Computing Group; Lau, B.Y.; Hoad, G.L.; Winsor, G.L.; Brinkman, F.S.L. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* **2017**, *45*, W30–W35. [CrossRef]
63. Bueris, V.; Sircili, M.P.; Taddei, C.R.; Santos, M.F.; Franzolin, M.R.; Martinez, M.B.; Ferrer, S.R.; Barreto, M.L.; Trabulsi, L.R. Detection of diarrheagenic *Escherichia coli* from children with and without diarrhea in Salvador, Bahia, Brazil. *Mem. Inst. Oswaldo Cruz* **2007**, *102*, 839–844. [CrossRef]
64. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
65. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]
66. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesleret, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
67. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]
68. Wattam, A.R.; Davis, J.J.; Assaf, R.; Boisvert, S.; Brettin, T.; Bun, C.; Conrad, N.; Dietrich, E.M.; Disz, T.; Gabbardet, J.L.; et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* **2017**, *45*, D535–D542. [CrossRef]
69. Dereeper, A.; Audic, S.; Claverie, J.; Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* **2010**, *10*, 8. [CrossRef]
70. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *36*, W465–W469. [CrossRef]
71. Chevenet, F.; Brun, C.; Bañuls, A.; Jacq, B.; Christen, R. TreeDyn: Towards dynamic graphics and annotations for analyses of trees. *BMC Bioinform.* **2006**, *7*, 439. [CrossRef]