CHAPTER 3

Identification and molecular characterisation of an Indian genotype of poliovirus type 1 isolated during outbreaks of poliomyelitis in the Democratic Republic of Congo (DRC) from 2006-2008.

3.1 INTRODUCTION

Poliomyelitis is caused by one of the wild type poliovirus serotypes 1, 2 and 3 and the great progress has been made by the Global Polio Eradication Initiative (GPEI) in interrupting these strains worldwide (CDC. 2006b by using three doses of live oral polio vaccine (OPV) in children under one year. This approach has stopped transmission of indigenous wild poliovirus (WPV) in all but 4 countries, Nigeria, India, Pakistan and Afghanistan (CDC. 2009a). The last case of WPV type 2 was detected in India in 1999 (CDC. 2001b). Unfortunately due to low vaccination coverage, cases of poliomyelitis continue to be reported in other regions as results of importations from endemic countries (van Niekerk, Vries et al. 1994) or emerging of vaccine derived polioviruses (VDPVs). OPV has been associated with adverse events as a results it causes vaccine associated paralytic polio (VAPP).

Laboratory surveillance has played an important role to provide to provide rapid information about the circulating WPV. Molecular tools have been developed and

implemented to help identify the serotype of the poliovirus and further perform molecular characterisation of wild strains. These tools have help to identify the presence of wild poliovirus infections and characteristic patterns of their distribution and spread (Chezzi, Blackburn et al. 1997). As described by Jorba et al., that the rate of poliovirus substitution is estimated at approximately 1% per year (Jorba, Campagnoli et al. 2008), this has given rise to genotype associated with specific regions(Chezzi, Blackburn et al. 1997). Molecular analysis has help to answer questions associated with location of endemic virus reservoirs, patterns of virus transmission or source of imported poliovirus.

Over 8500 AFP cases of both poliovirus type 1 and 3 were reported in Africa from January 2000 until 2008. Although most cases were concentrated in Nigeria, importation has resulted in outbreaks in a number of neighbouring and distant countries in West, East and Central Africa (CDC. 2002d).

In 2006, thirteen wild poliovirus1 cases and 3 contacts were reported in DRC. In 2007 this number increased to 41 before dropping back to 5 cases in 2008. The routine vaccine coverage in the DRC was estimated at 50% in 2008 (CDC. 2000a). This was even lower in regions that had reported wild poliovirus outbreaks.

We report the genetic characterization of the wild poliovirus 1 cases isolated during the period 2006-2008 in the DRC. The outbreak in the DRC originated from Angola, as suggested by genetic relatedness to wild poliovirus that circulated in Angola beginning in 2005. The Angola outbreak was due to imported virus of Indian genotype (SOAS)

(CDC.2006a; CDC. 2006d). This is the first time that AFP cases associated with the Indian genotype were identified in the DRC. This finding emphasises the vulnerability of regions with suboptimal vaccination coverage to the reintroduction of wild poliovirus from the remaining endemic countries.

3.2 MATERIAL AND METHODS

3.2.1 Viruses

Isolates from original stool specimens of suspected poliovirus AFP cases were submitted to the National Institute for Communicable Diseases (NICD) in South Africa for diagnostic reverse transcription polymerase chain reaction (RT-PCR), ELISA and partial genomic sequencing as described in Chapter 2. Virological investigation identified polioviruses of type 1.

3.2.2 Laboratory Diagnosis

Refer to Chapter 2

3.2.2.1 *Diagnostic RT-PCR*

Refer to Chapter 2

3.2.2.2 *ELISA*

Refer to Chapter 2

3.2.2.3 *RNA Extraction*

Refer to Chapter 2

3.2.2.4 *RT-PCR for sequence analysis*

Refer to Chapter 2

3.2.2.5 *Sequence Analysis*

Refer to Chapter 2

3.2.2.6 Phylogenetic Analysis

Phylogenetic analysis was carried out on the complete VP1 gene, which corresponds to 906 nt for all isolates from patients with AFP. Sequences were aligned with Clustal X (Thompson, Gibson et al. 1997) and analyzed using the software package Geneious [http://www.geneious.com]. Phylogenetic trees were inferred using Neighbour-Joining (Tamura, Dudley et al. 2007) , Maximum likelihood (Guindon, Lethiec et al. 2005) and Bayesian (Drummond and Rambaut 2007) methods as implemented in Geneious. Genetic distances were estimated under models of evolution correcting for multiple substitutions at a site and for unequal transition and transversion rates. Statistical support at each node of the inferred trees was assessed using common procedures for each method. Estimation of the evolutionary rate of VP1 sequences was calculated using Bayesian Markov chain Monte Carlo (MCMC) as implemented in BEAST v1.6 (Drummond and Rambaut 2007). Onset dates were used for estimating the evolutionary rate under the strict clock model (Drummond and Rambaut 2007). Briefly, two independent chains consisting of 40,000,000 steps each were run under the GTR+gamma model of evolution and the Coalescent Bayesian Skyline model as a tree prior. The GTR + G substitution model was used to allow different sites in the alignment to evolve at different rates and the substitution rate calculated from the data assuming a molecular clock. The Coalescent Bayesian Skyline model was used to estimate the posterior distribution of the population size through time. Sample correlation was assessed checking the Effective Size Samples (ESS) statistic. A maximum credibility tree was inferred by BEAST in which branch lengths were scaled to time according to

the estimated mean substitution rate. The annotated tree file was visualised with Figtree version 1.3.1 (http://tree.bio.ed.ac. uk/software/figtree).(Drummond and Rambaut 2007).

3.3 RESULTS

3.3.1 Outbreak description:

In the Democratic Republic of the Congo the first case of wild poliovirus was confirmed in Boma district in Bas Congo (BCG) province, at the border with Angola in February 2006. This case had a date of onset of paralysis on the 27th of February 2006 and was followed by another with an onset on the 24th of April 2006, from Sekebanza district in the same province of DRC. By the end of 2006, a total of 13 cases and 3 contacts had been reported in DRC from nine districts spread in four provinces. Most of these cases were from Kasai Occidental (KOC) province. Following the index case in the Boma district, no further circulation was reported from this district.

In 2007 additional 41 wild poliovirus cases plus 5 contacts were reported. Équateur (EQT) province reported more than 50% of these cases, whereas Orientale (ORT) province had 13 wild poliovirus cases, followed by Bandundu (BDD) province with 5 cases. Orientale province was not affected in 2006, but Bandundu province's outbreak continued until March 2007. In total, 17 districts were affected. The first wild poliovirus case in 2007 had an onset of paralysis on the 1st of January 2007 whereas the last case confirmed the paralysis on the 20th of November 2007. The spread of the wild poliovirus in DRC followed the traditional transport route of the river Congo. All poliovirus cases reported in 2006 and 2007 were confirmed as wild poliovirus 1.

In 2008, only 5 wild poliovirus PV1 cases were reported, affecting 4 provinces and 5 districts. In total, 58 AFP cases of wild poliovirus 1 were reported in the DRC during period of 2006 to 2008.

3.3.2 Prevalence and properties of the India strain in Africa:

Viral isolates obtained from DRC cases from 2006, 2007 and 2008 were tested by RT-PCR using pan-enterovirus, pan poliovirus, serotype specific, and Sabin type 1, 2 and 3 virus specific primers. All isolates were identified as wild poliovirus1. The RT-PCR results were also confirmed by ELISA. These isolates were characterised by sequencing of 906 bases comprising the complete VP1 gene.

3.3.3 Phylogenetic analysis:

Since all previous poliomyelitis cases in the DRC were associated with the African genotype WEAF-A, the genetic variability of all isolates identified in the outbreaks in the DRC was investigated by complete VP1 gene sequence analysis and included in the phylogenetic trees.

Blast search analysis of all polioviruses isolated in the current investigation identified them as SOAS (Indian) genotype. Comparison to reference strains of all wild poliovirus 1 genotypes confirmed the classification of the DRC strains to the SOAS-A genotype. Indian strains that were demonstrated to be the closest to the described outbreaks were included in the analysis. All polioviruses investigated fell within the SOAS-A genotype within the B2D1B genetic cluster (Table 3.1).

Table 3.1: Laboratory data of wild poliovirus strains detected in acute flaccid paralysis (AFP) from the DRC. Epid Number indicates Epidemiological number (EPID Number).

EPID Number	Onset date	Percentage	Best Match	Accession Numbers
RDC-SKV-MTM-08-001	2008.217	98.68	RDC-ORT-KIS-07-100	GU951803
RDC-SKV-KLH-08-003	2008.161	99.45	RDC-EQT-BOE-07-006	GU951802
RDC-NKV-RUT-08-002	2008.282	99.12	RDC-EQT-BOE-07-006	-
RDC-ORT-LUB-08-004	2008.45	99.45	RDC-EQT-BOE-07-006	GU951800
RDC-ORT-BAN-07-019	2007.324	99.23	RDC-ORT-KIS-07-094	GU951799
RDC-ORT-BAS-07-027	2007.286	99.23	RDC-EQT-BOE-07-006	GU951798
RDC-ORT-YAK-07-008	2007.287	99.34	RDC-ORT-BAS-07-017	-
RDC-ORT-KIS-07-100	2007.278	99.89	RDC-ORT-KIS-07-094	-
RDC-ORT-KIS-07-098	2007.273	99.56	RDC-ORT-KIS-07-094	-
RDC-ORT-BAS-07-023	2007.269	99.34	RDC-ORT-YAH-07-003	-
RDC-ORT-BAS-07-020	2007.242	99.45	RDC-EQT-BOE-07-001	-
RDC-ORT-BAS-07-019	2007.261	99.45	RDC-EQT-BOE-07-001	-
RDC-ORT-KIS-07-094	2007.267	99.12	RDC-EQT-BOE-07-001	-
RDC-ORT-BAS-07-017	2007.252	99.78	RDC-ORT-YAH-07-003	GU951797
RDC-ORT-BAS-07-016	2007.237	99.45	RDC-ORT-BAS-07-016	GU951796
RDC-ORT-YAH-07-003	2007.219	99.22	RDC-ORT-YAH-07-003	GU951795
RDC-EQT-BIK-07-013	2007.200	98.90	RDC-EQT-MBD-06-013	GU951794
RDC-EQT-BIN-07-009	2007.233	99.44	RDC-EQT-BOM-07-004	GU951793
RDC-BDD-BOL-07-006	2007.173	99.22	RDC-EQT-BOE-07-001	-
RDC-EQT-PIM-07-003	2007.143	99.67	RDC-EQT-BOE-07-001	-
RDC-EQT-PIM-07-002	2007.178	99.12	RDC-EQT-MBD-07-001	-
RDC-EQT-LIS-07-029	2007.157	98.90	RDC-EQT-BOE-07-001	-
RDC-EQT-LIS-07-030	2007.157	99.23	RDC-EQT-BOE-07-001	-
RDC-EQT-BOE-07-006	2007.138	99.78	RDC-EQT-BOE-07-001	-
RDC-EQT-BIK-07-009	2007.144	99.23	RDC-EQT-BOE-07-001	GU951792
RDC-EQT-ING-07-007	2007.102	98.90	RDC-BDD-INO-06-013	GU951791
RDC-EQT-MBD-07-018	2007.126	99.45	RDC-EQT-MBD-07-012	-
RDC-BDD-KIR-07-004	2007.67	99.12	RDC-BDD-INO-06-013	GU951790
RDC-BDD-KIR-07-003	2007.64	99.12	RDC-BDD-INO-07-002	GU951789
RDC-EQT-MBD-07-012	2007.53	99.45	RDC-EQT-BOE-07-001	-
RDC-EQT-BLB-07-002	2007.62	99.34	RDC-EQT-BOE-07-001	-
RDC-EQT-BLB-07-001	2007.44	99.45	RDC-EQT-MBD-07-001	-
RDC-BDD-KIR-07-001	2007.44	99.00	RDC-BDD-INO-06-013	-
RDC-BDD-INO-07-002	2007.10	99.89	RDC-BDD-INO-06-013	GU951788
RDC-KOC-LUE-06-004	2006.225	99.23	RDC-KOC-TKP-06-004	GU951785
RDC-BCG-MAT-06-001	2006.170	99.01	RDC-BCG-SEK-06-004	GU951784
RDC-KOC-TKP-06-006	2006.187	99.01	ANG-LSL-SAU-05-002	GU951783
RDC-KOC-KAO-06-005	2006.161	98.45	ANG-BEN-LOB-05-003	GU951782
RDC-EQT-LUK-07-005	2007.173	99.22	RDC-EQT-BOE-07-001	-
RDC-KOC-KAM-06-005	2006.140	98.34	ANG-LSL-SAU-05-002	GU951781
RDC-KOC-TKP-06-004	2006.125	98.34	ANG05-2006706767	GU951780
RDC-BCG-SEK-06-004	2006.114	99.56	RDC-BCG-BOM-06-001	GU951779
RDC-BCG-BOM-06-001	2006.58	98.68	ANG-BEN-LOB-05-003	GU951778
ANG-BEN-LOB-05-003	2005.139	99.45	ANG-LUA-CAC-05-003	EU046204

Phylogenetic relationships among 58 DRC sequences and 3 ANG related sequences are summarized in a Bayesian MCMC tree using complete VP1 (906 nt) sequences (Figure 3.1). The geographical and temporal clustering of the sequences in the tree suggests local circulation of multiple DRC lineages after importation of ANG strains. The topology of the tree is consistent with at least three distinct geographic lineages (A, B, and C). Lineage A groups sequences mainly from BCG and BDD provinces and it contains the first case dated in early 2006. The majority of cases belonged to lineage B comprising viruses found in EQT and ORT provinces. Two clearly defined groups defined lineage B; one group with sequences from EQT and the other group with cases mainly from EQT and ORT provinces. Viruses circulating exclusively in the KOC province during 2006 grouped in lineage C.

Figure 3.1 shows the Bayesian tree with an estimated root splitting the three lineages, including the related Angola strains, demonstrating the evolution that has occurred since the introduction of this genotype into Africa. Branch lengths from the tree were estimated under the strict clock model and displayed in time units (Figure 3.1). The inferred evolutionary rate was 1.11×10^{-2} substitutions per site per year (s/s/y) with estimated 95% highest posterior densities (95% HPD) ranging from 0.89×10^{-2} to 1.37×10^{-2} s/s/y. The estimated rate for the DRC lineages is consistent with rates estimated for other wild poliovirus genotypes and vaccine-derived poliovirus outbreaks (Jorba, Campagnoli et al. 2008). The estimated age of the most common recent ancestor to DRC and related ANG sequences was 3.54 years (95% HPD ranging from 3.32 to 3.70 years), approximately one year before the first DRC case and approximately 3 months before the first related ANG sequence. The topology of the tree is consistent with three main

lineages that spread widely over DRC. However, local multiple chains of transmission are inferred from each lineage. Long branches and/or low branch support (posterior values) characterize the three lineages. Date estimates inferred in internal nodes showed wide confidence intervals (not shown) because of gaps in phylogenetic signal indicative of gaps in surveillance.

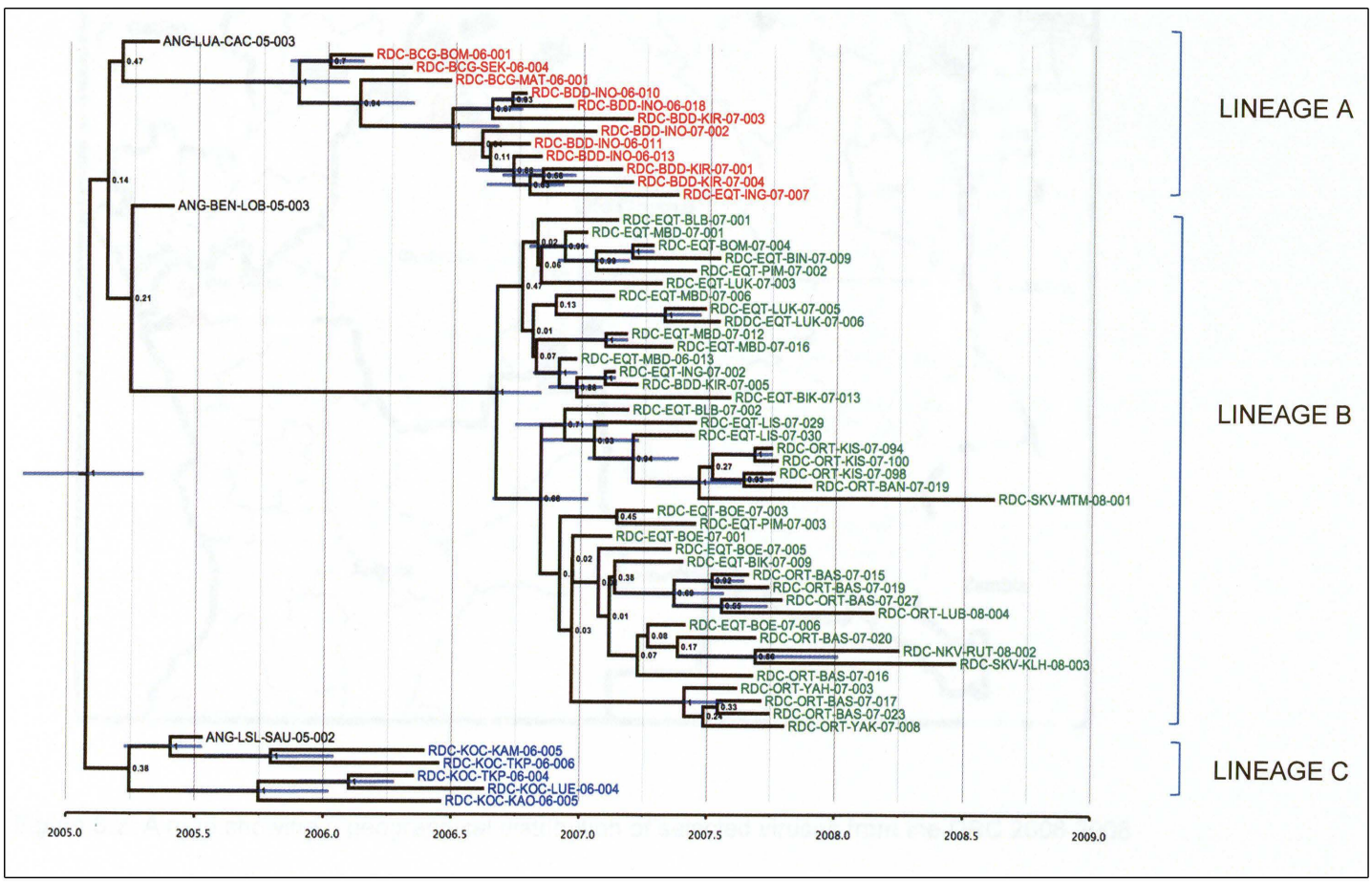

Figure 3.1: A Bayesian tree with an estimated root demonstrating the evolution over time. The scale indicate the time of first isolation (2005) to the last (2008) at a rate of 1% per year as demonstrated for poliovirus.

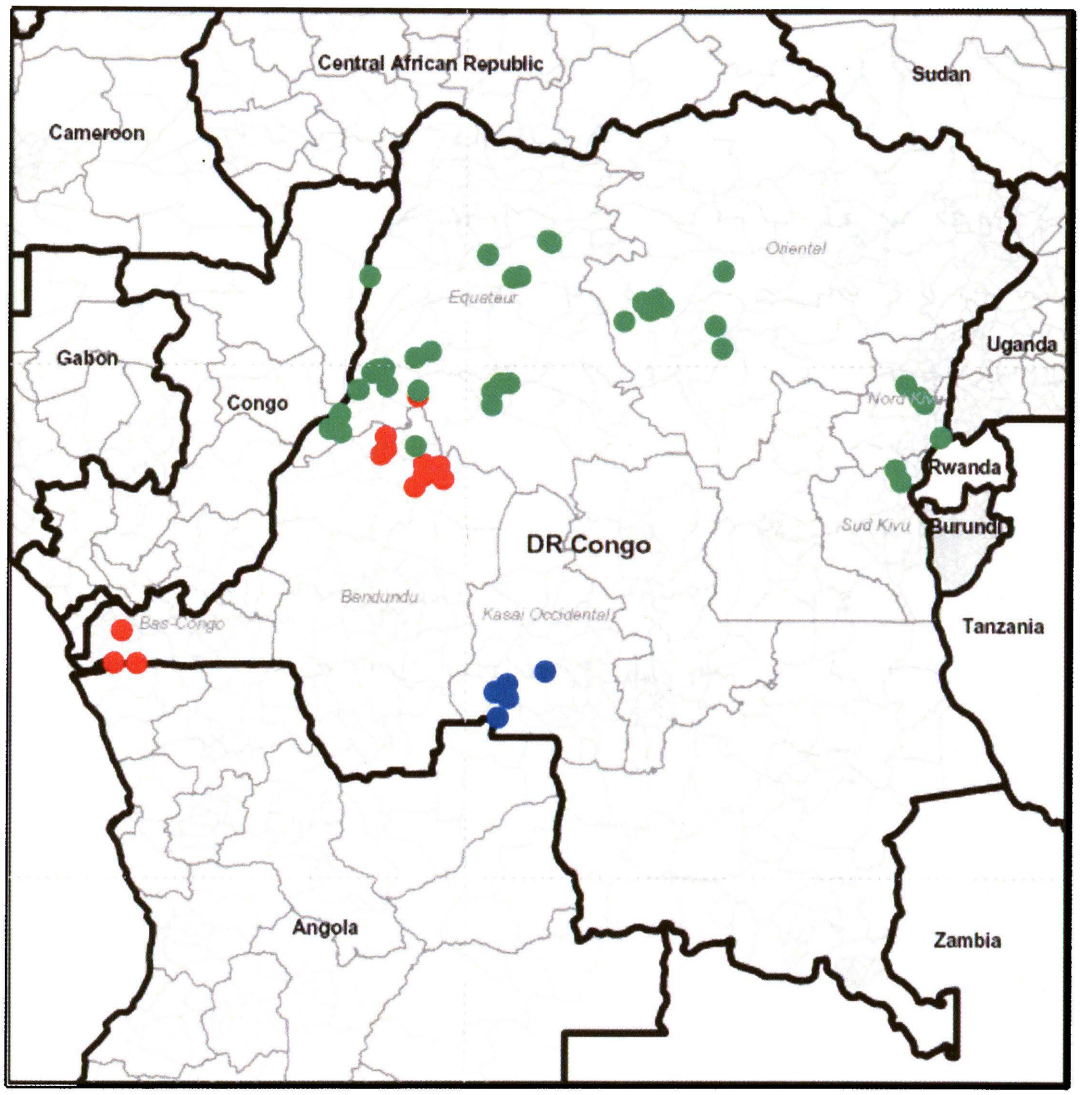

Figure 3.2: A map showing a geographical distribution of selected viruses from the DRC 2006-2008 outbreak. Red dots, lineage A, Green dots, lineage B and Blue dots lineage C.

Mapping and visualization of genetic lineages (Figure 3.2) was consistent with phylogenetic results. Separate importations from ANG into southern DRC (KOC and

BCG provinces) were detected during 2006, at least half a year after their closest ANG sequence. While cases in the KOC province circulated within the province (lineage C), lineage A spread north and east into the BDD and EQT provinces during more than 6 months. Interestingly, lineage C had higher nucleotide diversity (0.023±0.004 substitutions per site, n=5) than lineage A (0.011±0.002 substitutions per site, n=12) and lineage B (0.013±0.001 substitutions per site, n=41). The earliest isolates in lineage B clustered in the EQT province. Further inference about timing and source of importation events in lineage B were undetermined due to lack of sequence record during at least one year. Lineage B spread north and east covering an extended geographical range reaching in 2008 communities bordering with Uganda, Rwanda, and Burundi. In addition, local circulation across the interprovincial borders of EQT and BDD were inferred and visualized in both the tree and map.

3.4 DISCUSSION

In this study the complete VP1 gene of wild PV1 polio strains causing poliomyelitis during outbreaks in the DRC was sequenced to investigate the molecular epidemiology. The complete VP1 region has since been chosen for molecular epidemiological investigations as it exhibits the greatest sequence variability, and plays a key role in receptor attachment. (Wimmer, Hellen et al. 1993).

The first case associated with wild poliovirus was identified in February 2006 in the Boma district of the DRC. The Bayesian tree (Figure 3.1) which assumes a molecular clock demonstrates the drift that had occurred from the introduction of the Angola strains in the DRC in 2006 and strongly suggests that Angolan strains were the source for the DRC outbreaks. Most isolates from DRC were from the border provinces between the DRC and Angola, namely Equateur, Bandundu and Kasai Occidental. The DRC viruses from these border regions form separate virus lineages to those viruses circulating in Angola.

Wild poliovirus type 1 of SOAS genotype was first identified in the DRC in 2006. Three to four possible independent poliovirus lineages were identified, and they could be grouped according to their geographical location in the country. Viruses from Bas Congo and Bandundu formed lineage A, which were detected from 2006 to 2007. Kasai Occidental (KOC) viruses formed lineage B, which circulated in 2006 and was not detected thereafter. The viruses from the Equateur (EQT) province and viruses

from ORT formed lineage C. These results demonstrate widespread circulation and sustained transmission of individual strains along separate pathways.

Surveillance of AFP has been a key factor in efforts to achieve global eradication of poliomyelitis worldwide. The outbreaks in the DRC demonstrated that low levels of vaccine coverage, in combination with poor surveillance of AFP, could result in undetected and prolonged circulation of poliovirus in a community before cases of poliomyelitis are evident (Kew, Morris-Glasgow et al. 2002).

In South America an endemic circulating genotype was displaced by an important genotype due to low vaccine coverage (Rico-Hesse, Pallansch et al. 1987). In Africa, the WEAF-A genotype which was last detected in September 2000 in DRC was followed by the imported SOAS genotype in 2006. The DRC had been free of polio for the 6 years prior to the 2006 outbreak. The Indian genotype has been circulating for years and same strain has been reported in Bangladesh, (CDC. 2002c) and additional importations were reported in Nepal and Myanmar (CDC.1999).

As mentioned previously in Chapter 2, importations had underscored the global polio eradication goal and this can only be achieved by the elimination of remaining reservoirs of wild poliovirus endemicity in South Asia and Sub-Saharan Africa (CID. 2002).

In conclusion, this study describes the distribution and molecular epidemiology of wild PV1 SOAS genotype in southern and central Africa. The SOAS strains identified in the DRC were unique to Africa and are estimated to have circulated about 1 year in the

DRC or Angola before being detected. The DRC outbreak was caused by the SOAS genotype, but was introduced from neighbouring Angola, most likely through frequent cross-border movement among population groups living on both sides of the border. On the African continent, so far the SOAS genotype appears to be limited to certain countries in Southern and Central regions although virus circulation is still ongoing.