Comparative genomic analysis of malaria mosquito vector associated novel pathogen *Elizabethkingia anophelis*

Jeanette Teo[1¶], Sean Yang-Yi Tan[2,3¶], Yang Liu[2], Martin Tay[2,3], Yichen Ding[2], Yingying Li[3], Staffan Kjelleberg[2,4] , Michael Givskov[2,5], Raymond T.P. Lin[1], Liang Yang[2,3 *]

Table S1. General characteristics of the seven *Elizabethkingia spp.* strains from this study: three patient isolates (NUHP1, NUHP2 and NUHP3) and four sink isolates (NUH1, NUH4, NUH6 and NUH11). Values obtained from the RAST annotation server(Aziz, et al. 2008).

	*Elizabethkingia spp.*strains						
Characteristics	NUHP1	NUHP2	NUHP3	NUH1	NUH4	NUH6	NUH11
Genome							
Size (bp)	4,337,802	4,334,646	4,334,105	4,334,661	4,239,486	4,123,800	4,091,484
No. of contigs	65	59	71	59	50	74	59
GC content (%)	35.6	35.6	35.5	35.6	35.6	35.6	35.6
No. of coding sequences	4054	4046	4063	4060	3971	3870	3826
No. of subsystems	330	329	329	329	329	327	328
No. of RNAs	48	46	47	48	48	48	50

Table S2. General characteristics of *E. anophelis* strains Ag1 and R26, and *E. meningoseptica* NBRC 12535. Values obtained from the RAST annotation server (Aziz, et al. 2008).

	E. anophelis	*E. anophelis*	*E. meningoseptica*	*E. meningoseptica*	*E. meningoseptica*
Characteristics	Ag1	R26	ATCC13253(NITE)	ATCC13253(OSU)	502
Genome					
Size (bp)	4,045,712	4,032,720	3,840,286	3,796,928	3,960,658
No. of contigs	51	66	34	115	21
GC content (%)	35.5	35.4	36.4	36.6	35.5
No. of coding sequences	3825	3759	3520	3487	3652
No. of subsystems	323	321	323	323	330
No. of RNAs	43	40	41	48	52

Table S3. Proteins involved in virulence encoded from core- and accessory-genomes of the *Elizabethkingia spp.* identified by BLAST search against the Virulence Factors of Pathogenic bacteria Database (VFDB).

Database id	**% Identity**	**Annotation**
core-genome		
VFG0028	32.98	(gi:33594368) brkB - serum resistance protein
VFG0036	44.79	(gi:33591349) bplC - lipopolysaccharide biosynthesis protein
VFG0070	30.32	(gi:16802625) iap - P60 extracellular protein, invasion associated protein Iap
VFG0073	31.82	(gi:16802247) plcA - phosphatidylinositol-specific phospholipase c
VFG0077	52.08	(gi:16804506) clpP - ATP-dependent Clp protease proteolytic subunit
VFG0079	45.71	(gi:16802278) clpC - endopeptidase Clp ATP-binding chain C
VFG0080	49.27	(gi:16803037) clpE - ATP-dependent protease
VFG0082	59.96	(gi:15640836) aldA - aldehyde dehydrogenase
VFG0087	43.83	(gi:15640841) tagD - tagD protein
VFG0101	36.51	(gi:15640855) toxT - TCP pilus virulence regulatory protein
VFG0119	39.82	(gi:15600454) algR - alginate biosynthesis regulatory protein AlgR
VFG0120	33.85	(gi:15600455) algZ - sigma factor AlgU
VFG0122	33.7	(gi:15598736) algD - GDP-mannose 6-dehydrogenase AlgD
VFG0130	31.66	(gi:15598744) algI - alginate o-acetyltransferase AlgI
VFG0152	49.18	(gi:15598673) rhlR - transcriptional regulator RhlR
VFG0160	31.61	(gi:15597593) pvdE - pyoverdine biosynthesis protein PvdE
VFG0167	30.41	(gi:15599423) pchR - transcriptional regulator PchR
VFG0168	34.04	(gi:15599424) pchD - pyochelin biosynthesis protein PchD
VFG0307	30.22	(gi:15644871) napA - neutrophil activating protein (napA) (bacterioferritin)

VFG0313	38.96	(gi:15645788) gluP - glucose/galactose transporter (gluP)
VFG0314	32.22	(gi:15644988) gluE - UDP-glucose 4-epimerase
VFG0320	44.3	(gi:15646084) kdtB - lipopolysaccharide core biosynthesis protein (kdtB)
VFG0321	32.73	(gi:15644675) rfbD - GDP-D-mannose dehydratase (rfbD)
VFG0323	30.93	(gi:15645486) lpxB - lipid A disaccharide synthetase (lpxB)
VFG0328	37.31	(gi:16273439) lic1C - lic-1 operon protein (licC)
VFG0344	30.96	(gi:16272073) hitC - iron(III) ABC transporter, ATP-binding protein
VFG0358	32.39	(gi:16122154) fyuA/psn - pesticin/yersiniabactin receptor protein
VFG0366	39.5	(gi:16122162) ybtQ - inner membrane ABC-transporter YbtQ
VFG0428	31.03	(gi:16763118) tviE - Vi polysaccharide biosynthesis protein TviE, Glycosyl transferases group 1
VFG0431	53.61	(gi:16763123) tviB - Vi polysaccharide biosynthesis protein, UDP-glucose/GDP-mannose dehydrogenase
VFG0478	30.88	(gi:16764063) fur - transcriptional repressor of iron-responsive genes (Fur family) (ferric uptake regulator)
VFG0479	39.87	(gi:16764728) pykF - pyruvate kinase I (formerly F), fructose stimulated
VFG0483	34.9	(gi:16764732) orf408 - putative regulatory protein, deoR family
VFG0487	32.97	(gi:16764736) ttrS - Tetrathionate reductase complex: sensory transduction histidine kinase
VFG0491	31.88	(gi:16764740) orf242 - putative regulatory proteins, merR family
VFG0492	30.05	(gi:16764741) ssrB - Secretion system regulator: transcriptonal activator, homologous with degU/uvrY/bvgA
VFG0555	30.28	(gi:16766200) invC - surface presentation of antigens; secretory proteins
VFG0562	37.76	(gi:16766215) mutS - methyl-directed mismatch repair, recognize exocyclic adducts of guanosine
VFG0574	53.53	(gi:16767047) mgtB - Mg<up>2+</up> transport protein

VFG0575	40.54	(gi:16767048) mgtC - Mg<up>2+</up> transport protein
VFG0576	42.57	(gi:16767506) ssb - ssDNA-binding protein controls activity of RecBCD nuclease
VFG0585	33.33	(gi:16767515) soxS - transcriptional activator of superoxide response regulon (AraC/XylS family)
VFG0596	42.67	(gi:16764454) copR - Copper resistance; transcriptional regulatory protein
VFG0670	39.54	(gi:24111748) gtrB - bactoprenol glucosyl transferase
VFG0679	32.88	(gi:6470206) dep/capD - <gamma>-glutamyltranspeptidase
VFG0696	44.64	(gi:30962644) bexA - ATP-dependent polysaccharide export protein
VFG0841	35.33	(gi:3822163) hlyB - hemolysin transport protein
VFG0864	31.76	(gi:471302) aggR - AraC homolgous regulator of AAF/I
VFG0869	35.38	(gi:34148232) aatC - AatC ATB binding protein of ABC transporter
VFG0872	32.21	(gi:26251200) fimE - Type 1 fimbriae Regulatory protein fimE
VFG0916	34.23	(gi:26250129) chuS - Putative heme/hemoglobin transport protein
VFG0917	30.23	(gi:26250130) chuA - Outer membrane heme/hemoglobin receptor
VFG0922	37.82	(gi:26250139) chuU - Putative permease of iron compound ABC transport system
VFG0923	30.81	(gi:26246561) fepA - Ferrienterobactin receptor precursor
VFG0925	31.8	(gi:26246567) fepC - Ferric enterobactin transport ATP-binding protein fepC
VFG0934	33.33	(gi:26246575) entA - 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase
VFG0937	31.75	(gi:26249459) iucD - IucD protein
VFG1036	34.25	(gi:15808716) tetA(B) - tetracycline resistance protein TetA(B)
VFG1037	39.22	(gi:15808717) tetC - putative transcriptional regulator TetC
VFG1038	31.37	(gi:15808718) tetD - putative transcriptional regulator TetD
VFG1050	30.81	(gi:15808730) orf34 - unknown
VFG1066	42.59	(gi:15808746) orf50 - unknown

VFG1085	41.43	(gi:14626628) iutA - IutA
VFG1105	37.88	(gi:15641775) VC1772 - hypothetical protein
VFG1115	30.65	(gi:15641785) nanK - ROK family protein
VFG1206	36.56	(gi:15677997) fbpC - iron(III) ABC transporter, ATP-binding protein
VFG1214	30.1	(gi:15599743) pilR - two-component response regulator PilR
VFG1225	37.31	(gi:15595605) pilG - twitching motility protein PilG
VFG1226	33.33	(gi:15595606) pilH - twitching motility protein PilH
VFG1234	35.64	(gi:15595613) chpD - probable transcriptional regulator
VFG1248	49.35	(gi:15596294) fleQ - transcriptional regulator FleQ
VFG1266	34.38	(gi:15599418) pchI - probable ATP-binding component of ABC transporter
VFG1269	33.81	(gi:33591935) cyaB - cyclolysin secretion ATP-binding protein
VFG1300	39.96	(gi:21281856) cap8D - capsular polysaccharide synthesis enzyme Cap8D
VFG1306	45.45	(gi:21281862) cap8J - capsular polysaccharide synthesis enzyme Cap8J
VFG1342	38.1	(gi:22537323) cpsO - glycosyl transferase CpsO(V)
VFG1349	30.66	(gi:22537330) cpsD - cpsD protein
VFG1354	40.82	(gi:15900810) cbpE - choline binding protein E
VFG1381	60.38	(gi:15607608) icl/aceA - aceA
VFG1390	36.77	(gi:15608121) mprA - hypothetical protein Rv0981
VFG1394	38.25	(gi:15608893) plcD - plcD
VFG1395	34.29	(gi:15608948) mgtC - mgtC
VFG1398	32.89	(gi:15609329) trpD - trpD
VFG1404	34.36	(gi:15609565) ahpC - ahpC
VFG1406	35.34	(gi:15609848) ideR - ideR
VFG1408	30.08	(gi:15610077) mas - mas

VFG1411	43.55	(gi:15610124) leuD - leuD
VFG1412	31.82	(gi:15610359) sigH - sigH
VFG1416	57.01	(gi:15610737) panD - panD
VFG1417	40.7	(gi:15610738) panC - panC
VFG1444	30.99	(gi:9965749) aslA - putative arylsulfatase
VFG1449	43.28	(gi:146567) kpsT - KpsT
VFG1486	34.07	(gi:24528017) ORF35 - putative CS12 fimbrial-like upstream regulatory protein
VFG1557	35.56	(gi:23954241) hlyB - HlyB protein
VFG1584	32.12	(gi:23954267) orf50 - hypothetical protein
VFG1587	45.05	(gi:23954270) orf53 - hypothetical protein
VFG1633	31.1	(gi:14594867) mchF - microcin transport protein MchF
VFG1650	30.14	(gi:14594880) iroN - IroN protein
VFG1653	30.66	(gi:28316233) iroC - ABC transport protein
VFG1661	31.07	(gi:28316241) orf38 - hypothetical protein
VFG1668	45.96	(gi:28316248) orf45 - putative lysil-tRNA synthetase LysU
VFG1669	37.14	(gi:28316249) orf46 - hypothetical protein
VFG1722	30.08	(gi:26249436) c3601 - Hypothetical protein
VFG1747	31.39	(gi:16120593) YPO0256 - two-component sensor/regulator
VFG1751	30.33	(gi:16120597) YPO0260 - putative AraC-family regulatory protein
VFG1794	34.75	(gi:15088598) ysrS - YsrS
VFG1797	36.75	(gi:25011354) scpB - streptococcal C5a peptidase
VFG1826	32.27	(gi:15609720) relA - relA
VFG1855	65.31	(gi:52840925) htpB - Hsp60, 60K heat shock protein HtpB
VFG1859	32.65	(gi:52842863) feoB - ferrous iron transporter B

VFG1861	63.33	(gi:52840449) katA - catalase/(hydro)peroxidase
VFG1864	34.87	(gi:52841028) mip - macrophage infectivity potentiator (Mip)
VFG1866	42.13	(gi:52841515) rpoS - stationary phase specific sigma factor RpoS
VFG1867	43.01	(gi:52843161) sodB - superoxide dismutase
VFG1889	31.4	(gi:52842852) letA - response regulator GacA
VFG1931	34.58	(gi:15792793) cadF - outer membrane fibronectin-binding protein
VFG1939	35.37	(gi:15792463) Cj1138 - putative galactosyltransferase
VFG1958	30.12	(gi:15792748) Cj1430c - putative nucleotide-sugar epimerase/dehydratase
VFG1965	30.23	(gi:15792755) Cj1437c - putative aminotransferase
VFG1971	44.01	(gi:15792761) kpsF - KpsF protein
VFG1974	55.56	(gi:15792764) kpsT - putative capsule polysaccharide export ATP-binding protein
VFG2008	36.43	(gi:52842007) motB - flagellar motor protein MotB
VFG2059	41.3	(gi:15595271) PA0073 - probable ATP-binding component of ABC transporter
VFG2064	42.67	(gi:15595276) PA0078 - hypothetical protein
VFG2085	45.22	(gi:15600888) vasH - sigma-54 dependent transcriptional regulator
VFG2110	36.96	(gi:52843027) vipD - VipB interferes with multivesicular body formation at the late endosome
VFG2222	33.09	(gi:17987678) manC - mannose-1-phosphate guanylyltransferase
VFG2228	44.64	(gi:17987699) wzt - O-antigen export system ATP-binding protein
VFG2230	33.33	(gi:17987701) wbkC - GDP-mannose 4,6-dehydratase / GDP-4-amino-4,6-dideoxy-D-mannose formyltransferase
VFG2276	45.65	(gi:18309155) colA - collagenase
VFG2279	33.33	(gi:18310216) nagJ - hyaluronidase
VFG2281	33.63	(gi:18310505) nagL - hyaluronidase

VFG2284	30.6	(gi:18309535) nanJ - exo-alpha-sialidase
VFG2361	45.95	(gi:123443279) galE - UDP-glucose 4-epimerase
VFG2363	32.47	(gi:123443281) manC - mannose-1-phosphate guanylyltransferase
VFG2374	30.52	(gi:123443292) prt - paratose synthase
VFG2376	30.63	(gi:123443294) ddhB - CDP-glucose 4,6-dehydratase
VFG2440	33.33	(gi:53722543) bprB - two-component response regulator
VFG2445	30.67	(gi:53722548) bapB - acyl carrier protein
VFG2480	45.52	(gi:53722524) clpV - Clp-type ATPase chaperone protein
VFG2526	30.63	(gi:53720911) cheB - chemotaxis-specific methylesterase
VFG2532	36.75	(gi:53720917) cheY1 - chemotaxis two-component response regulator CheY1
VFG2538	38.1	(gi:53721911) pmlR/bspR1 - N-acylhomoserine lactone dependent regulatory protein
VFG2546	31.74	(gi:53720396) wcbT - acyl-CoA transferase
VFG2550	33.94	(gi:53720400) wcbP - capsular polysaccharide biosynthesis dehydrogenase/reductase
VFG2563	39.06	(gi:53720413) wzt2 - ATP-binding ABC transporter capsular polysaccharide export protein
accessory-genome		
VFG0032	51.79	(gi:33591345) bplG - probable sugar transferase
VFG0033	32.31	(gi:33591346) bplF - lipopolysaccharide biosynthesis protein
VFG0036	32.6	(gi:33591349) bplC - lipopolysaccharide biosynthesis protein
VFG0037	40.79	(gi:33591350) bplB - probable acetyltransferase
VFG0167	35.37	(gi:15599423) pchR - transcriptional regulator PchR
VFG0208	33.01	(gi:15596910) exsA - transcriptional regulator ExsA
VFG0315	39.2	(gi:15644674) rfbM - mannose-6-phosphate isomerase

VFG0364	32.1	(gi:16122160) ybtA - transcriptional regulator YbtA
VFG0365	35.92	(gi:16122161) ybtP - lipoprotein inner membrane ABC-transporter
VFG0428	30.92	(gi:16763118) tviE - Vi polysaccharide biosynthesis protein TviE, Glycosyl transferases group 1
VFG0564	33.33	(gi:16767037) sugR - ATP binding protein
VFG0574	33.54	(gi:16767047) mgtB - Mg<up>2+</up> transport protein
VFG0596	40	(gi:16764454) copR - Copper resistance; transcriptional regulatory protein
VFG0660	41.18	(gi:24114260) yeeS - intergenic-region protein
VFG0780	30.36	(gi:11527916) orf7 - unknown
VFG0864	36.73	(gi:471302) aggR - AraC homolgous regulator of AAF/I
VFG0869	33.33	(gi:34148232) aatC - AatC ATB binding protein of ABC transporter
VFG0917	33.6	(gi:26250130) chuA - Outer membrane heme/hemoglobin receptor
VFG0934	34.12	(gi:26246575) entA - 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase
VFG1028	30.39	(gi:15808708) intI1 - Tn21 integrase IntI1
VFG1030	30.69	(gi:15808710) tnpR - resolvase TnpR
VFG1038	35.87	(gi:15808718) tetD - putative transcriptional regulator TetD
VFG1066	35.85	(gi:15808746) orf50 - unknown
VFG1074	31.97	(gi:15808754) orf58 - IS1328 transposase
VFG1085	35.71	(gi:14626628) iutA - IutA
VFG1100	36.26	(gi:15641770) VC1767 - conserved hypothetical protein
VFG1234	30.38	(gi:15595613) chpD - probable transcriptional regulator
VFG1267	32.6	(gi:15599419) pchH - probable ATP-binding component of ABC transporter
VFG1301	63.82	(gi:21281857) cap8E - capsular polysaccharide synthesis enzyme Cap8E
VFG1341	30.91	(gi:22537322) cpsJ - glycosyl transferase CpsJ(V)

VFG1342	36.76	(gi:22537323) cpsO - glycosyl transferase CpsO(V)
VFG1344	39.91	(gi:22537325) cpsM - polysaccharide biosynthesis protein CpsM(V)
VFG1348	41.5	(gi:22537329) cpsE - glycosyl transferase CpsE
VFG1372	37.33	(gi:15900282) cps4H - capsular polysaccharide biosynthesis protein Cps4H
VFG1388	36.27	(gi:15608039) ompA - ompA
VFG1389	31.07	(gi:15608043) prrA - hypothetical protein Rv0903c
VFG1456	31.75	(gi:24527988) ORF5 - hypothetical protein
VFG1473	32.8	(gi:24528005) ORF22 - hypothetical protein
VFG1486	35.09	(gi:24528017) ORF35 - putative CS12 fimbrial-like upstream regulatory protein
VFG1511	57.72	(gi:24528040) ORF60 - putative integrase
VFG1512	33.16	(gi:24528041) ORF61 - putative reverse transcriptase
VFG1539	30.85	(gi:23954223) hek - adhesin/virulence factor Hek
VFG1569	48.89	(gi:23954253) orf35 - hypothetical protein
VFG1583	31.52	(gi:23954266) orf49 - hypothetical protein
VFG1633	66.67	(gi:14594867) mchF - microcin transport protein MchF
VFG1650	30.34	(gi:14594880) iroN - IroN protein
VFG1706	38.62	(gi:26249410) c3575 - Transposase insF for insertion sequence IS3A/B/C/D/E/fA
VFG1731	31.33	(gi:26249445) c3610 - Putative receptor
VFG1746	42.59	(gi:16120592) YPO0255 - putative two-component response regulator
VFG1922	31.52	(gi:15792655) ptmA - putative oxidoreductase (flagellin modification)
VFG1931	41.59	(gi:15792793) cadF - outer membrane fibronectin-binding protein
VFG1936	38.49	(gi:15792460) Cj1135 - putative two-domain glycosyltransferase
VFG1937	37.21	(gi:15792461) Cj1136 - putative galactosyltransferase
VFG1938	34.17	(gi:15792462) Cj1137c - hypothetcal protein Cj1137c

VFG1939	39.33	(gi:15792463) Cj1138 - putative galactosyltransferase
VFG1947	38.61	(gi:15792471) waaV - putative glucosyltransferase
VFG1956	47.31	(gi:15792746) fcl - putative fucose synthetase
VFG1968	37.5	(gi:15792758) Cj1440c - putative sugar transferase
VFG2027	33.96	(gi:15792474) waaE - putative ADP-heptose synthase
VFG2044	36.67	(gi:33592926) bvgA - virulence factors transcription regulator
VFG2225	67.14	(gi:17987696) gmd - GDP-mannose 4,6-dehydratase
VFG2226	35.5	(gi:17987697) per - perosamine synthetase
VFG2228	39.15	(gi:17987699) wzt - O-antigen export system ATP-binding protein
VFG2284	30.6	(gi:18309535) nanJ - exo-alpha-sialidase
VFG2343	33.55	(gi:123442779) flgJ - flagellar protein FlgJ
VFG2366	30.16	(gi:123443284) rfpB - putative galactosyltransferase
VFG2368	35.14	(gi:123443286) wbcG - putative glycosyltransferase
VFG2373	31.97	(gi:123443291) wbcC - putative glycosyltransferase
VFG2464	37.21	(gi:53722567) bsaN - AraC family transcriptional regulator
VFG2546	31.59	(gi:53720396) wcbT - acyl-CoA transferase