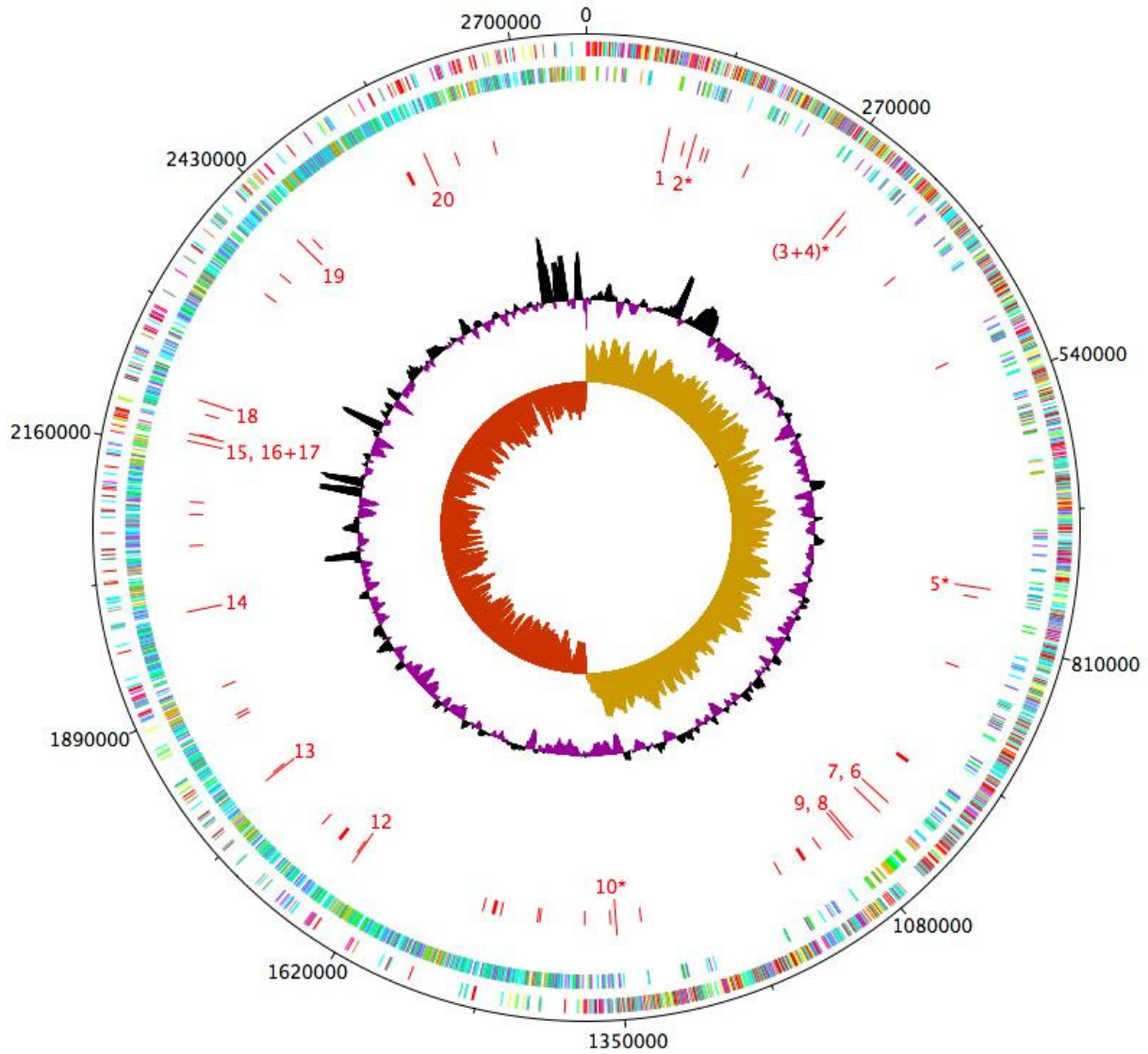


Helgenomisk karaktärisering av isolat från fågelbotulismutbrott. (H0743012)**Projektledare/huvudman: Bo Segerman**

En översikt översiktsbild över generna i genomet från ett svenskt fågelbotulismisolat

Hypoteser/Frågeställningar:

- Genom att förstå den genetiska sammansättningen av *Clostridium botulinum* (*C. botulinum*) kan man bättre förstå patogenisitetsmekanismerna bakom fågelbotulism och därmed få ett underlag till hur forskning och åtgärder bör prioriteras.

*Vi har nu fått en mycket klarare bild av de komplexa mekanismer som ligger bakom fågelbotulism. Dessa involverar ett samspel mellan flera komponenter: plasmider, bakteriofager, toxingener samt mobila DNA element som alla finns i flera upplagor och som varierar mellan olika isolat. Alla dessa komponenter finns i en genomisk bakgrund som påminner mycket om *Clostridium novyi*. Detta projekts färdigställande utgör inte bara ett slut utan starten på nya möjligheter att angripa fågelbotulism vilket diskuteras i slutet av rapporten.*

- Jämförande studier bör även kunna ge värdefull vägledning till vilken grad resultat från forskning kring human botulism kan översättas till fågelbotulism.

Det står nu klart att bakterien som orsakar fågelbotulism är mycket skild från de humana typerna. Om man bortser från just toxingenen skulle dessa inte klassificeras i samma art.

- Hur ser den genetiska bakgrunden ut i de *C. botulinum* stammar som just nu förekommer i Sverige och finns det någon skillnad från tidigare insamlade isolat?

De stammar som figurerar just nu skiljer sig en hel del, både genomiskt, i plasmiduppsättningen, i bakteriofagerna och de mobila elementen jämfört med äldre svenska isolat och andra labstammar som det finns information.

- Kan detta hjälpa oss att finna genetiska orsaker till varför fågelbotulism plötsligt ökat så kraftigt?

De relativt stora skillnaderna mellan de nya kliniska isolaten och äldre isolat tyder på att en ny variant har uppstått som är bättre anpassad till fåglar. Bakterien verkar ha en mycket stor möjlighet att variera sitt genetiska material, speciellt med avseende på plasmider, fager och mobila element. Vi kan identifiera vissa gemensamma nämnare bland de nya isolaten men riktade typningsmetoder för dessa bör utvecklas.

- Kan tillgång till sekvenser från alla bakteriens ytproteiner möjliggöra vaccinutveckling?

Vi har nu tillgång till alla proteiner och annoteringar. Det finns stora möjligheter att välja ut vaccinkandidater från dessa. Detta kan vara ytproteiner, toxiner, fagrelaterade proteiner, plasmidrelaterade proteiner eller faktorer som påverkar koloniseringsmöjligheten. Åtskilliga sådana proteiner har identifierats.

- Kan tillgången till hela sekvensen möjliggöra utveckling av bättre och effektivare PCR diagnostik?

Vi har nu en stor uppsättning primrar som spänner över hela genomet. Vi har nu möjlighet att köra reaktioner mot specifika kromosomala, plasmid, fag och toxin mål. Vi har alltså möjlighet att typa isolat med avseende på ett stort antal faktorer.

- Kan den kunskap om odlingsbetingelser och upprening av isolaten som planeras förvärvas användas för bättre diagnostik och möjliggöra nya forskningsprojekt?

Att kunna odla och isolera bakterien är en förutsättning för att effektiv forskning och typning skall kunna ske. Många laboratorier runtom i världen är eniga om att det är en mycket svårödlad bakterie. Detta var ett av de hinder vi slogs med i början av projektet och det har vi lyckats övervinna.

- Kan kunskapen om den genetiska sammansättningen av svenska isolat ligga till grund för att vidta förebyggande åtgärder för hur sjukdomen introduceras och sprids i fjäderfä?

Vi kan konstatera att det vilda isolat vi studerat skiljer sig med avseende på plasmiduppsättning, mönster av mobila element samt toxingenernas lokaliseringar. De två tamdjursisolaten (från olika år) är dock helt lika i dessa aspekter. Detta är en stark indikation på att utbrotten hos tamfågel och vildfågel inte är direkt kopplade till varandra. Typning på ett större antal isolat baserat på denna genominformation krävs dock för att vara helt säker. Den genomiska informationen öppnar möjligheter att testa hypoteser kring hur man skulle kunna skapa instabilitet i fag/plasmidpopulationen genom yttre faktorer (ett fenomen som vi slåss mot för att hålla stammarna intakta på lab).

Bakgrund:

Fågelbotulism beror på förgiftning med botulinum-neurotoxin (BoNT) som produceras av bakterier inom gruppen *C. botulinum*. Toxinet är ett nervgift som förlamar fåglarna. En kraftigt ökad mängd utbrott har de senaste åren konstaterats i svenska fjäderfäbesättningar. I detta projekt utnyttjas den nya parallellsekvenseringsteknologin för att kartlägga genomen på de stammar som orsakar fågelbotulism i Sverige idag.

C. botulinum finns i 7 typer benämnda A-G efter deras toxintyp samt i 4 ”grupper” baserat på övergripande egenskaper (tabel 1).

Tabell 1: Indelning av *Clostridium botulinum* i typ och grupp, patogenicitet i fåglar vs. Människa samt neurotoxigenens placering.

C. BOTULINUM FENOTYPISKA GRUPPER



Fågelbotulism orsakas framförallt av typ C (eller, som vi numera har kommit fram till, en typ som är till en del C och en del D) men även typ A, D och E kan gå på fåglar. Typ C och D bär sina botulinumtoxigener på bakteriofager, typ G på en plasmid och övriga på kromosomen.

Material och metoder:

Isolering: Bakterierna växtes vid 37°C i 9 ml pre-reduced TPGY broth (5% Tryptone [Difco], 0.5% Proteose Peptone [Difco], 0.4% glucose, 2% Yeast Extract [Oxoid] and 0.1% starch [Merck]) supplementerat med 0.1% L-cystein-HCl and 0.14% NaHCO).

DNA preparation: DNA preparerades med Quiagen DNeasy. Dåligt utbyte var normalt men polning av flera preparationer gav tillräcklig mängd DNA för genomsekvensering.

Genom sekvensering och assembly: Genomen för "Clin1" (gödkyckling 2007), "Stockholm" (SVA isolat från tidigt 50 tal), "CB3" (Ett vildfågelsisolat) samt "CB4" (gödkyckling 2008) sekvenserades med Roche 454 teknologi. Sekvenserna "assemblades" med Roche newbler assembler program. Mellan 300-400 contiger (pusselbitar att sättas ihop) skapades.

Slutning av genomet: att sätta ihop pusselbitarna ("contiger") och fylla ut saknade regioner ("gaps"). En del gaps löstes genom att manuellt inspektera och jämföra ändarna med hjälp av programmen "Consed och Mummer. PCR primrar designades mot övriga gaps och PCR produkter sekvenserades. De gaps som ändå inte gick att lösa annoterade vi ändarna på och lyckades på så sätt minska sätten de kunde kombineras. På slutet var det de mobila elementen som återstod att para ihop med varandra. Analys av insersionssekvenserna underlättade detta arbete.

Annotering: Annotering skedde med Blastgämförelser. Programmet Artemis användes för att underlätta processen. Annoteringar från närbesläktade genom (*C.novyi*) överfördes via Mysql databasprogram. Gener som inte kunde annoteras via närbesläktade arter analyserades manuellt.

Övrig genomanalys: Förutom de ovannämnda programmen utvecklades ett flertal specialdesignade datorverktyg för att underlätta analyserna i detta projekt.

Resultat:

Steg1: Isolering:

Efter initiala svårigheter i isoleringssteget har vi lyckats komma fram till en metod som fungerar för isolering och odling. Följande isolat har tagits fram:

Tabell2: Isolat från fågelbotulismutbrott isolerade i projektets initiala fas.

Isolat	Utbrottsår	Fågel	Plats	Bakteriofag
BKT 07 22828	2006	Tam	*****	Tappad
BKT 07 28387 ("clin1")	2007	Tam	*****	Tappad
07 V891 ("CB3")	2007	Gråtrut	Vollholmen	+
BKT 07 28396	2007	Tam	*****	Tappad
BKT 07 29909	2007	Tam	*****	Tappad
BKT 08 08340	2007	Tam	*****	+
BKT 08 35155	2008	Tam	*****	+
V 07881	2007	Gråtrut	EKÅ	(+)
V 07877	2007	Gråtrut	EKÅ	(+)
C6N07	2007	Tam	*****	+
Hättorp 08 09258	2008	Tam	*****	+
Skepparp 3:	2007	Tam	*****	+
BKT 08 35533	2008	Tam	*****	(+)
BKT 08 15925 ("CB4")	2008	Tam	*****	+
Vinninga 09 69465	2008	Tam	*****	+
Sävare Lunneberg (Vinninga)	2008	Tam	*****	+

För att underlätta det praktiska datorarbetet har arbetsnamnen Clin1 = 07-BKT028387, CB3 = 07-V891 and CB4 = 08-BKT015925 för de stammar som användes för genomsekvensering.

Steg2: Provbearbetning

Fågelbotulismbakterien visade sig ge dåligt utbyte i många genom-DNA-preparationsmetoder. Detta var ytterligare en svårighet vi fick överbrygga. Flera olika metoder utvärderades (fenolkloroformbaserade, "charge switch" baserade, silicabaserade). Poolning av flera DNeasy (Quiagen) preparationer visade sig bäst. Det var också viktigt att gå in med mer DNA än standardnivåerna vid 454 genomsekvenseringen då förluster fanns i förbehandlingsstegen.

Steg3: Parallellsekvensering:

Vi skickade det första isolatet vi fått fram ("Clin1") till parallellsekvensering. Bakteriofagen som bär på toxingenen hade vid tidpunkten för sekvenseringen sjunkit i nivå så pass mycket så att det inte fanns tillräckligt med sekvensdata för att sätta ihop den. Vi valde då att gå tillbaks och jobba på isolering och odlingsförhållanden ytterligare. Vi fick sedermera fram isolat som var bättre och två av dessa (CB3 och CB4) skickades på nytt till 454 parallellsekvensering. Resultatet var nu mycket gott och det fanns en stor mängd sekvensdata av mycket god kvalitet. I projektet har även sekvensdata för SVAs lab-stam "Stockholmsstammen" producerats.

Tabell3, summering av parallellsekvenseringsdata:

	Datum	454 kemi	Antal sekvenser	Medellängd	Mängd sekvensdata (baser)	Täckning (medel)
Clin1	2008-10-10	FLX	260322	257	66965645	22.3
Stockholm	2008-10-10	FLX	103776	228	23610423	7.9
CB3	2009-10-22	Titanium	385592	252	96978118	32.3
CB4	2009-11-24	Titanium	853266	288	245793046	81.9

Steg 4: Hopsättning till draftsekvens

Sekvensdata som producerats i steg 3 sattes ihop ("assemblades") till draftsekvenser (WGS) med hjälp av programmet newbler (denovo assembly). Resultatet blev följande:

	Contigs (pusselbitar)	Approx. Genomstorlek	GC innehåll (%)
Clin1	264	2838894	28.0
Stockholm	443	2610208	27.4
CB3	323	3130325	27.9
CB4	545	2868528	27.6

Steg5: Hoppusslande av draftsekvensens delar (contiger) till en komplett sekvens

Eftersom CB4 hade mest sekvenser och var från slaktkyckling valde vi att fokusera på detta isolat. CB4 genomet drevs alltså till komplett stadium. Övriga genom kunde genom jämförelser och referensguidad assembly drivas till nästan komplett stadium. Vi fokuserade framförallt på att få ihop plasmiderna korrekt för dessa samt lokalisera mobila elementens insersionsställen.

Tyvärr var CB4 assemblyn inte den bästa (fler pusselbitar än de andra). Genom att använda de andra assemblyna som referens lyckades vi dock ganska snart få ner CB4 till samma nivå som de övriga. Vi satte ihop ett stort antal contig med hjälp av manuell analys av ändarna. Sedan designade och körde vi PCR över troliga hopsättningar. Slutligen annoterade vi ändarna som var kvar och konstaterade att det var "mobila element" som var mest svårlösta", ett slags DNA element som "hoppas runt i genomet" och kan finnas i många nästan helt lika upplagor. Genom att assemblera sekvenserna mot consensus av dessa, köra långa PCR reaktioner, jämföra isolat samt insersionssekvenser lyckades vi lösa samtliga gaps under sommaren 2010. CB4 genomet innehöll 1 kromosom, 4 plasmider och en stor bakteriofagplasmid (tabell 5).

Tabell 5: CB4 genomets delar

CB4	storlek (kilobaser)	GC innehåll (%)
kromosom	2760	28.5
Bakteriofag	202.6	26.5
plasmid 90kb	89.8	26.1
plasmid 78kb	77.7	26.7
plasmid 40kb	40.1	28.0
plasmid 11kb	11.1	26.3

Steg 6: Slutpolering av sekvensen

Slutpolering innebär att små "stavfel" i den nästan 3 miljoner tecken lång sekvensen rättas till. Detta är en process som man kan jobba mycket med. Vi har rättat till de flesta småfelen med automatiserade metoder men även gått igenom stora mängder sekvens manuellt. Framförallt plasmiderna. Vi jobbar kontinuerligt med att söka efter eventuellt ytterligare småfel så att den är så perfekt som möjligt när den publiceras i Genbank.

Steg 7: Definiering och annotering av gener

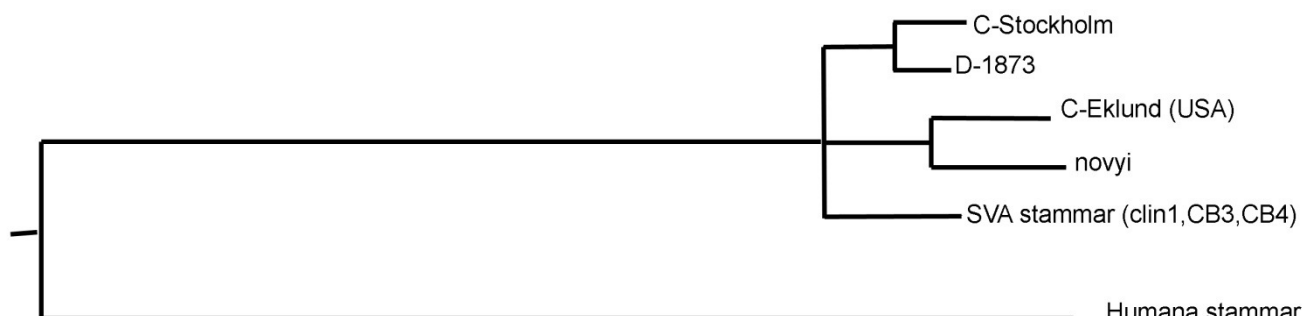
Generna definierades med hjälp av gensökningsprogrammet "Glimmer" Vi hittade totalt 3140 gener (tabell7). Fördelningen mellan kromosomen och plasmider/fager visas också. Generna var mycket starkt klustrade på + strängen på ena halvan av genomet och på minussträngen på andra halvan (syns tydligt på figuren på framsidan). Detta mönster följde även GC avvikelsen (innersta cirkeln på framsidefiguren). Även plasmiderna hade tendens att ansamlas gener på samma sträng. Detta tyder på att transkriptionen är selekterad för optimal effektivitet. Bakterien är troligen utsatt för hård konkurrens i sin nisch.

Tabell 7 Gener.

	storlek	Gener	Medel gentorlek	gentäthet (gener/kb)	rRNA	Mobila element	pro-fag
Kromosom	2760398	2648	898	0.96	10	59	Nej
Fag 200kb	202582	229	713	1.1	0	5	Ja
Plasmid 90kb	89829	94	723	1.0	0	5	Nej
Plasmid 80kb	77720	97	611	1.2	0	7	Nej
Plasmid 40kb	39640	59	549	1.5	0	0	Ja
Plasmid 11kb	12403	13	578	1.0	0	0	Ja?
Totalt	3182572	3140	862.5	0.99	10	76	

Steg 8: jämförande genomik

Vi kan konstatera att både gödkycklingsisolaten och det vilda isolatet är mycket nära besläktat baserat på total genuppsättning. De äldre isolaten (Stockholm samt data från USA) är mer avlägset besläktade. Avståndet är ungefär som till *C. novyi*. Trots att de nya stammarna är mycket lika i "kärnan av genomet" visar det sig att de har olika uppsättning av plasmider samt olika mönster av mobila element, fager och toxingener. Det vilda isolatet har en annan uppsättning plasmider samt ett annat mönster av mobila element än isolaten från gödkyckling som var mycket lika. Avståndet till de humana varianterna av *Clostridium botulinum* är tillräckligt stort för att det borde klassas som andra arter. Däremot borde *C. Botulinum* typ C och D klassas som en variant av *C.novyi* med en speciell plasmid/faguppsättning.



Steg 9: Analys av plasmider, fager, mobila element, toxingener mm.

Genomet visade sig vara fullt med mobila element. I tabell 8 visas en översikt av dem 76 element vi identifierat. Vi har även närmare studerat insertionsmönster och sekvenser. Vi har kommit fram till att ISBma2 Sätter sig i hårnålsstrukturer som motsvarar transkriptionsstoppsignaler i genomet. På så vis

förstörs inte genfunktioner. Det verkar finnas ett element (primärelement) som driver de övriga att hoppa runt (sekundärelement). Elementet kan således transkriptionsstopsignalsmönstret och omforma operon. ISCPe7 tycks snarare känna igen strukturer före generna. INT visade sig ofta sätta sig innuti RT/RT. Sammantaget har vi kommit fram till att genomet har ett aktivt och mycket komplext system av mobila element som troligen påverkar genomets plasticitet och anpassningsmöjligheter.

Tabell 8: Mobila element i genomet.

	Kromosom	Fag 200kb	Plasmid 90kb	Plasmid 80kb	Plasmid 40kb	Plasmid 11kb
ISCPe7	23	0	1	2	0	0
ISBma2	17	0	0	1	0	0
Tpmut	3	1	0	1	0	0
INT/IST	6	2	3	2	0	0
RT/RT	10	0	0	1	0	0
TPtk	0	2	1	0	0	0
totalt	59	5	5	7	0	0

Plasmiduppsättningen i CB4 och Clin1 var samma men CB3 och stockholmsstammen hade en annorlunda plasmiduppsättning. (se tabell 6). Ytterligare två andra plasmider verkar finnas i D1875 stammen efter studier i genbank. Plasmiderna i olika stammar har segment som är gemensamma och det verkar således finnas ett genetiskt utbyte mellan olika plasmider.

Tabell 6: Plasmiduppsättning i de olika isolaten

plasmid/fag	CB4	Clin1	CB3	Stockholm
Bakteriofag (200kb)	x	x	x	
Bakteriofag (180kb)				x
plasmid 90kb	x	x	x	
plasmid 78kb	x	x	x	
plasmid 50kb			x	x
plasmid 40kb	x	x	x	
plasmid 11kb	x	x		
plasmid 12kb				x

Det är känt att botulinumtoxigenen ligger på en fag. I våra stammar är denna fag större och bitvis helt olik den fag som finns i stockholmsstammen. Fagen är en eller möjligen ligger i en plasmid och deltar sålunda troligen i det genetiska utbytet. Även 40kb plasmiden kodar för en komplett bakteriofag men har även plasmidreglerande gener. 11kb plasmiden har också fag-relaterade gener. Ytterligare fagrelaterade gener kan komma att dyka upp vid fördjupad analys av genomet.

200 kb/180kb fagen bär på botulinumtoxigenen. I CB3 och CB4 ligger en C2 toxigen i 80kb plasmiden medan i stockholmsstammen finns det i 12kb plasmiden. Vi har även hittat ett nytt toxin, epsilon-toxin typ B i 11 kb plasmiden. Det finns även en hel del andra gener i 80 och 90 kb plasmiderna som potentiellt kan ha med kolonisation, patogenicitet och resistens att göra.

Diskussion

Vi har nu en fullständig bild av *Clostridium botulinum* typ C (C/D) genomet. Det finns en mycket konserverad kärna i genomet och sedan finns ett antal system baserat på plasmider, fager och mobila element som aktivt omformar genomets egenskaper. Denna plasticitet ger genomet en förmåga att anpassa sig till sin omgivning men gör också att till exempel toxigener kan gå förlorade vid kultivering i laboratorium. Detta projekts avslutning innebär egentligen början på helt nya möjligheter att ta sig an fågelbotulismproblemet. Jag hoppas i framtida projekt kunna utnyttja denna kunskap till att skapa mer

riktade typningsmetoder för att kunna läsa av plasmidtyper, fagtyper, toxinotyper osv. Vi har på SVA betydligt mer material och får även in nytt och analys av dessa kommer förhoppningsvis ge en klarare bild av vilka komponenter/faktorer som har betydelse. Generna för dessa faktorer kan utnyttjas i vaccinforskningsprogram. Vi har även möjlighet att söka stimuli som rubbar plasticitetsbalansen i genomet som därmed kan förlora sin toxicitet. Det skulle till exempel kunna vara en aktivering av mobila element som stör replikeringen av den stora toxinbärande fagen.

Publicering/Resultatförmedling

Detta projekt fick av flera anledningar (problem med isolering, DNA preparation och kötider för sekvensering) en trög start och uppskov med rapporteringen har sökts och beviljats. Vidare gjorde den stora förekomsten av mobila element det svårarbetat i pussellösarsteget. Projektet har gått för full fart hela vintern, våren och slutspurten har varit kontinuerligt under hela sommaren. Helhetsbilden har sålunda fallit på plats mycket sent och vi jobbar just nu på vetenskapligt och populärvetenskapligt manus som vi kommer att skicka in snarast möjligt. Vi hoppas även muntligen kunna presentera dessa data i så stor utsträckning som möjligt.