



UPPSALA  
UNIVERSITET

# CRISPR/cas9: funktion och bakteriofagernas samevolution

Carolina B. Viman

---

Independent Project in Biology  
Självständigt arbete i biologi, 15 hp, vårterminen 2017  
Institutionen för biologisk grundutbildning, Uppsala universitet

# CRISPR/cas9: funktion och bakteriofagernas samevolution

Carolina B. Viman

Självständigt arbete i biologi 2017

## Sammandrag

Bakteriofager kallas de virus som infekterar bakterier. Bakteriofagernas förmåga att replikera sig beror mycket på hur väl utrustad dessa är för att kunna kontrollera bakterien den har infekterat. Arkéer utsätts också för samma typ av hot från virus, alltså är det viktigt för alla prokaryoter (encelliga organismer utan cellkärnmembran) att utveckla ett inre skydd för att kunna skydda sig mot virus DNA som tar sig in i cellerna. Ett av dessa inre skydd som prokaryoter har lyckats utveckla är ett adaptivt immunförsvar kallat CRISPR/cas. CRISPR (clustered regularly interspaced short palindromic repeats) är ett lokus i prokaryoternas kromosomer. CRISPR loci har påträffats i ungefär 90 procent av arkéer och nära 50 procent av bakterier. För att en prokaryot ska kunna vara resistent mot ett virus krävs det att prokaryotens CRISPR lokus innehåller en eller flera sekvenser som är komplementära till sekvenser från virusets DNA. De insatta virus-sekvenserna kallas för "spacers", och dessa spacers kommer från sekvenser från virusets DNA som kallas för "protospacers".

Framför CRISPR lokus sitter *cas* generna. De kodar för proteiner som har olika uppgifter inom CRISPR/cas immunsystemet. I CRISPR typ II system är det proteinet cas9 som står för immuniteten, då det är det som klyver virus DNA:t och gör det oanvändbart. Cas9 proteinet i sig består av två lober som har olika uppgifter, en igenkänningslob och en nukleaslob. Dessa lober är i behov att två andra komponenter för att kunna fungera som de ska: tracrRNA och crRNA. Det är två typer av RNA som transkriberas från CRISPR lokus och har som uppgift att guida cas9 till rätt protospacer. Var i en protospacer cas9 klyver bestäms av PAM sekvensen. PAM är ett fragment som sitter två till tre baspar innan en protospacer och är en mycket konserverad sekvens. PAM sekvenserna som känns igen ser dock olika ut beroende på vilken bakterie cas9 härstammar från.

För att virus ska kunna fortsätta kunna sprida sitt DNA och föröka sig så behöver de ha någon form av samevolution med CRISPR/cas. Samevolutionen mellan bakteriofager och bakterier med avseende på CRISPR/cas9 ser mycket olika från bakterie till bakterie. De kan alla ha CRISPR/cas9 men sedan också andra typer av immunförsvar som varierar mellan dem. De evolutionära förändringar i en bakteriofag som motverkar CRISPR/cas9 måste därför även innehålla anpassningar till andra eventuella immunsystem runt om. Hur organismer samevolverar med immunförsvar kan ge oss mer kunskap om utvecklingen av vårt eget immunförsvar. Kanske kan vi i framtiden rikta evolutionen av vårt immunförsvar till att ge oss ett bättre skydd från bakterier och virus. I en tid som denna när vår överanvändning av antibiotika har lett till att vi inte längre kan bli botade från alla infektioner är det en upptäckt som kan vara mycket viktig för vår överlevnad.

## Inledning

Ett virus förmåga att replikera sig beror mycket på hur väl utrustad denne är för att kunna kontrollera prokaryoten den har infekterat. En bakteriofag kan vara utrustad med verktyg för att snabbare kunna replikera sig väl inne i sin värd, till exempel kan den bryta ned värdens DNA för att få komma först i kön för att få sitt egna DNA transkriberat (Bull *et al.* 2004). Under evolutionen har prokaryoter behövt skydda sig mot dessa typer av angrepp från virus. Prokaryoterna har till exempel utvecklat mer utrustade membran för att förhindra att viruset

tar sig in, men virusen har också utvecklat smartare strategier för att ändå kunna ta sig förbi prokaryoternas yttre skydd. Samevolutionen mellan prokaryoter och virus verkar gå extra fort fram när det handlar om gener kopplade till infektioner eller skydd mot infektioner (Paterson *et al.* 2010). Prokaryoterna har, förutom att utveckla yttre skydd, även behövt utveckla ett inre skydd för att kunna skydda sig mot virus DNA som tar sig in i cellerna. Ett av dessa inre skydd är ett adaptivt immunförsvar kallat CRISPR/cas där CRISPR (clustered regularly interspaced short palindromic repeats) är ett lokus i prokaryoternas genom och cas är proteiner som har olika funktioner i det adaptiva immunförsvaret. CRISPR/cas systemet sparar gensekvenser från invaderande virus DNA för att sedan använda dessa som ett immunologiskt minne som sitter i det egna genomet (Mojica *et al.* 2005).

CRISPR loci finns i ungefär 90 procent av alla arter av arkéer och nära 50 procent av bakterier (Mojica *et al.* 2000). Det är en region av prokaryoternas kromosomer som innehåller specifika repetitiva sekvenser som separeras med icke-repetitiva sekvenser. De repetitiva sekvenserna brukar vara ungefär 21-37 baspar långa medan de icke-repetitiva sekvenserna är 20 baspar och går under namnet "spacers" (Jansen *et al.* 2002). Framför CRISPR lokus sitter en ledarsekvens. Ledarsekvensen brukar innehålla en promotor och är betydligt längre än repetitiva sekvenser, oftast ett par hundar baspar lång. De har ofta en hög andel A och T nukleotider och även långa sträckor homopolymerer. Man har bara funnit dessa ledarsekvenser i närheten av CRISPR lokus, och ingen annan stans i genomet (Jansen *et al.* 2002).

Flera konserverade gener vars position i genomet precis gränsar till CRISPR loki är proteinkodande och mycket viktiga för att alla delar av CRISPR-cas immunförsvaret ska fungera. Dessa gener döptes till CRISPR-associerade gener, eller förkortat "cas gener". De brukar ofta sitta några hundra baspar framför eller bakom CRISPR lokus, beroende på vilken CRISPR typ de sitter i. I typ II är cas generna lokaliserade framför CRISPR lokus (Jansen *et al.* 2002). I typ II system sitter det ofta tre till fyra stycken cas gener tillsammans och dessa gener kodar för olika typer av cas proteiner (Jansen *et al.* 2002). Cas proteinerna bidrar till CRISPR/cas systemen på olika sätt. En av de mest studerade CRISPR/cas systemet är typ II CRISPR/Cas9 från bakterien *Streptococcus pyogenes*. I *S. pyogenes* innehåller CRISPR fyra olika cas proteiner (Figur 1.), däribland cas9 (Heler *et al.* 2015). Cas9 är ett nukleas som är mycket viktig för att skydda cellen mot invaderande virus DNA, då det är den som klyver DNA:t och gör det obrukbart (Garneau *et al.* 2010).

Framför cas generna i kromosomen sitter ett tracrRNA (trans-aktiverande CRISPR RNA). tracrRNA transkriberas och bidrar bland annat i bearbetningen av transkriberat CRISPR lokus (pre-crRNA) (Deltcheva *et al.* 2011).



Figur 1. I ett CRISPR typ II-A lokus från *S. pyogenes* visas repetitiva sekvenser i gult och icke-repetitiva sekvenser i grått. cas generna *cas9*, *cas1*, *cas2* och *csn2* visas med röda boxar. tracrRNA visas med gul pil och ledarsekvens med grå pil. Bilden är omarbetad från Karvelis *et al.* (2013).

Trots att så många prokaryoter har detta adaptiva immunsystem så händer det ändå att de blir

infekterade av virus. Det kan bero på att CRISPR/cas inte ger full resistens mot virus, men det kan också bero på att virus samevolverar med CRISPR/cas för att undkomma immunsystemet. Det krävs en djupare förståelse för hur CRISPR/cas systemet fungerar för att kunna föra en diskussion kring samevolutionen mellan virus och prokaryoter, speciellt med CRISPR/cas9 i fokus. Arbetet kommer här att gå igenom utseende och funktion av CRISPR loki och cas9 proteinet för att sedan föra en diskussion kring huruvida bakteriofager samevolverar med detta system och hur det i sådana fall skulle kunna gå till. Vad har CRISPR/cas9 för påverkan på samevolutionen mellan bakteriofager och bakterier?

## Vad är CRISPR/cas9?

### CRISPR lokus

För att en prokaryot ska kunna vara resistent mot ett virus krävs det att prokaryotens CRISPR lokus innehåller en eller flera sekvenser som är komplementära till sekvenser från virusets DNA (Mojica *et al.* 2005). Mojica *et al.* (2005) sekvenserade spacers från olika stammar av bakterier och arkéer och fann att de i många fall är identiska till DNA sekvenser från virus. De utförde även försök där de använde bakterier och arkéer innehållande komplementära spacers som de utsatte för de specifika virus som dessa spacers är komplementära till. Det visade att de bakterier och arkéer som innehöll en komplementär spacer inte kunde bli infekterad av det specifika viruset. En annan studie av Deveau *et al.* (2008) visar att *Streptococcus thermophilus* antar nya spacers i sitt CRISPR lokus efter att ha blivit utsatta för specifika bakteriofager. Dessa spacers visade sig efter sekvensering vara komplementära till en eller flera sekvenser från bakteriofagernas DNA. De upptäckte även att när nya spacers integreras i CRISPR lokus, så kommer äldre spacers att tas bort (Deveau *et al.* 2008). Dessa upptäckter tyder på att CRISPR tillhör ett adaptivt immunförsvar, och att detta immunförsvar dessutom är ärftligt då CRISPR är en del av cellens genom som replikeras och förs vidare till dotterceller.

En studie utförd av Garneau *et al.* (2010) med *S. thermophilus* har visat att en bakterie som har en spacer med en muterad nukleotid i 5'-ändan fortfarande kan vara resistent mot bakteriofager som har en liknande protospacer. Ett år senare publicerades också en studie av Manica *et al.* (2011) som visar att bakteriofager som har flera mutationer utspridda i sin protospacer kan infektera celler och att antalet infektioner ökar ju mindre komplementär den är till spacern i CRISPR. Det verkar finnas ett samband mellan hur många komplementära sekvenser ett CRISPR lokus innehåller och hur bra resistens en prokaryot har mot ett visst virus. Ju fler komplementära sekvenser ett CRISPR lokus innehåller, desto bättre resistens kommer prokaryoten att ha. Experiment som har utförts med bakterien *S. thermophilus* som man har utsatt för bakteriofager har visat att det inte verkar finnas något mönster för vad protospacer sekvenserna kodar för, de kan även bestå av en icke-kodande sekvens (Barrangou *et al.* 2007).

Det finns flera olika typer av CRISPR lokus som har olika uppgifter och befinner sig i olika typer av arter och organismer. De CRISPR lokus som är länkade till *cas9* gener kallas för CRISPR typ II. Cas9 finns i alla typ II system och har en viktig roll i att ge resistens mot bakteriofager (Barrangou *et al.* 2007). CRISPR typ II är uppdelat i tre olika grupper som är baserade på vilka *cas* gener dessa gränser till. De tre grupperna kallas för typ II-A, typ II-B och typ II-C (Chylinski *et al.* 2013). CRISPR typ II-A innehåller *cas* generna *cas9*, *cas1*, *cas2* och *csn2* (Figur 1.). Typ II-B innehåller samma gener förutom *csn2* som är utbytt mot *cas4*. I typ II-C finns det bara de tre första *cas* generna; *cas9*, *cas1* och *cas2*. CRISPR typ II har

hittills bara hittats i bakterier (Chylinski *et al.* 2013).

### **Cas9 proteinet**

Garneau *et al.* (2010) har visat att cas9 protein klyver bakteriofagers genom inom dess protospacer; en mycket viktig funktion för att åstadkomma resistens mot bakteriofager. Efter analys av klyvda DNA fragment från bakteriofager upptäckte de att klyvningen sker tre baspar efter dess PAM (protospacer adjacent motif). PAM är ett fragment som sitter två till tre baspar innan en protospacer och består av en mycket konserverad sekvens. *S. thermophilus* PAM består av 5'-NNAGAAW-3' (Garneau *et al.* 2010). Vad PAM har för sekvens och hur lång den är beror på ursprunget av det cas9 protein som känner igen sekvensen (Heler *et al.* 2015). Andra exempel på PAM:s är 5'-NGG-3' från *S. pyogenes* (Mojica *et al.* 2009), eller 5'-NNNNGATT-3' från *Neisseria meningitidis* (Hou *et al.* 2013).

Efter analyser av kristalliserade cas9 protein har Nishimasu *et al.* (2014) funnit att cas9 består av olika domäner, och tillsammans med studier av cas9s funktion har de kunnat dra slutsatser om vad dessa domäner har för uppgift. Nishimasu och hans team har studerat kristalliserat cas9 bundet till sgRNA (single guide RNA) och ett komplementärt DNA. Ett sgRNA är ett färdigbehandlat crRNA (CRISPR RNA) som är bundet till ett tracrRNA. I detta kristalliserade komplex fann de att cas9 bestod av två stora lober; en nukleaslob och en igenkänningslob (Nishimasu *et al.* 2014).

### *Nukleasloben*

Nukleasloben av cas9 består av tre olika domäner som alla medverkar i att se till att klyvningen av komplementärt virus DNA fungerar. De tre domänerna kallas RuvC, PI (PAM-interagerande) och HNH (Nishimasu *et al.* 2014).

RuvC är en mycket konserverad domän som har i uppgift att klyva den icke-komplementära strängen av virus DNA (Jinek *et al.* 2012). Strukturellt liknar den endonukleaset RNase H. Denna strukturella likhet säger även mycket om RuvCs funktion då den också är ett endonukleas som klyver den icke-komplementära strängen av DNA som binder in till cas9 (Nishimasu *et al.* 2014). Virus DNA:t binder till RuvC genom interaktioner mellan aminosyror på RuvC och kolatomer (position C2) på DNA:t (Nishimasu *et al.* 2014). RuvC är uppbyggt av tre motiv kallade I, II och III. Tillsammans med PI domänen bildar RuvC ett positivt laddat område av cas9 där negativt laddat sgRNA kan binda in med sin 3'-ände (Nishimasu *et al.* 2014).

Efter undersökning av PI domänens struktur och i jämförelse med andra proteiner har man sett att domänen är specifik för just cas9 proteinet. Inga andra liknande strukturer har hittats i andra proteiner (Nishimasu *et al.* 2014). PI domänen är lokaliserad perfekt för att kunna känna igen PAM på den icke-komplementära DNA strängen, och det är precis den uppgiften man tror att PI domänen har (Nishimasu *et al.* 2014). Virus DNA och PI domänen binder till varandra genom interaktion mellan aminosyror från PI och fosfatgrupper från DNA (Nishimasu *et al.* 2014). Försök har utförts på cas9 med borttagen PI domän för att se vad PI domänen kan ha för funktion. Resultaten visade att cas9 inte kan klyva virus DNA utan den. Detta resultat tyder på att PI domänen utgör en mycket viktig del av nukleaslobens funktion (Nishimasu *et al.* 2014). Jinek *et al.* (2012) utförde studier för att upptäcka just vad i en protospacer som PI domänen känner igen. Studien utfördes på cas9 från *S. pyogenes* och de upptäckte då att det är sekvensen 5'-NGG-3' som känns igen. Denna studie bevisade även att det måste vara just 5'-NGG-3', och inte den komplementära 3'-NCC-5', för att PI domänen ska kunna känna igen PAM sekvensen. Alltså är det den icke-komplementära DNA strängen av virus DNA:t som blir igenkänt (Jinek *et al.* 2012).

HNH domänen av cas9 har i uppgift att klyva den komplementära strängen av virus DNA (Jinek *et al.* 2012), där den klyver DNA:t tre nukleotider framför PAM (Nishimasu *et al.* 2014). HNH domänen sitter mellan motiv II och III av RuvC, men har annars mycket få bindningar som kopplar ihop det till resten av cas9 (Nishimasu *et al.* 2014). När Nishimasu och hans team undersökte kristalliserade cas9 proteiner fann de att HNH domänen bidrog till en stor skillnad mellan konformationerna på de olika cas9 proteinerna. Detta kan tyda på att HNH är en flexibel domän. Flexibiliteten kan bidra till att sgRNA och virus DNA lättare kan binda in till domänen (Nishimasu *et al.* 2014).

### *Igenkänningsloben*

Igenkänningslobens främsta uppgift är att binda sgRNA och virus DNA. PI domänen som bidrar till igenkänning av virus DNA kan man tro hör hemma i igenkänningsloben. Anledningen till att PI domänen tillhör nukleasloben beror på att den bidrar till nukleasaktiviteten av cas9, det gör inte några av de domäner som finns inom igenkänningsloben.

Bland typ II CRISPR/cas system är igenkänningsloben den minst konserverade delen av cas9, den skiljer sig mycket i storlek bland de tre olika cas9 typerna som finns (Nishimasu *et al.* 2014). Strukturellt har man inte hittat några likheter mellan igenkänningsloben och andra proteiner, och det verkar därför som att loben är specifik för just cas9. Loben är uppbyggd av tre domäner: REC1, REC2 och helixbron (Nishimasu *et al.* 2014).

REC1 har en stor påverkan på cas9:s funktion då den interagerar direkt med virus DNA. Fosfatgrupper och kolatomer i virus DNA:t binder till REC1. Det är detta som kopplar samman REC1 med PI och RuvC domänerna i nukleasloben, då virus DNA binder till PI med fosfatgrupper och till RuvC med kolatomer (Nishimasu *et al.* 2014).

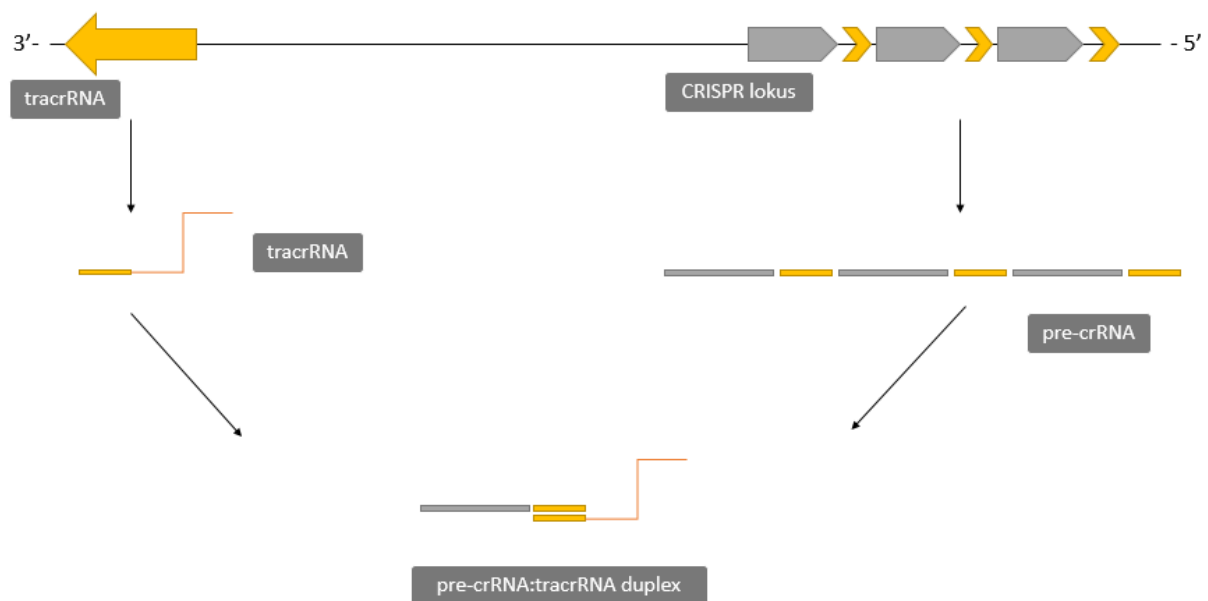
REC2 domänen är en del av cas9 vars uppgift ännu är oklar. Man har inte funnit någon bindning alls mellan REC2 domänen och sgRNA eller virus DNA, därför tror man att igenkänningsloben kan klara av relativt stora förändringar i REC2 domänen utan någon större förändring i funktionen hos cas9 (Nishimasu *et al.* 2014). Detta testade man genom att ta bort REC2 domänen och sedan kolla hur cas9 fungerade utan den. Resultatet var en cirka 50 procent lägre nukleas aktivitet hos cas9. Mutationer kan ha stor påverkan på uttrycket av gener, så Nishimasu antog att minskningen berodde på att mutationen orsakade lägre genuttryck av REC2 snarare än att domänen har så stor påverkan (Nishimasu *et al.* 2014).

Virus DNA:t binder även in till den tredje och sista domänen i igenkänningsloben: helixbron. DNA:t binder till helixbron genom sina fosfatgrupper. Mutationer i helixbron har visat att mängden klyvt virus DNA minskar, vilket tyder på att den har en viktig roll i cas9:s funktion. Studier har visat att helixbron interagerar med nukleotider från 3'-änden av virus DNA i närheten av PAM (Nishimasu *et al.* 2014). Denna region av ett virus DNA har döpts till PAM-proximala regionen och har en mycket viktig roll i klyvningen av en protospacer (Jinek *et al.* 2012). PAM-proximala regionen består av 10-12 nukleotider som sitter nära PAM i 3'-änden av protospacern. Studier har visat att om dessa nukleotider muteras kommer det leda till att klyvningen minskar eller stoppas helt. Samtidigt har man muterat nukleotider i 5'-änden av en protospacer och funnit att man kan mutera upp till sex nukleotider utan att ge en påverkan på klyvningen. PAM-proximala regionen är också viktigt för bindningen av sgRNA till protospacern (Jinek *et al.* 2012). I helixbron finns det även en konserverad region med aminosyran arginin. Denna region finns inom alla CRISPR typ II system och är mycket viktig för funktionen, då mutationer i den har visats leda till minskad aktivitet hos cas9 (Jinek *et al.*

2012).

### tracrRNA och crRNA

Tang *et al.* (2002) undersökte icke-kodande RNA från arkéen *Archaeoglobus fulgidus*. De fann då att vissa av dessa RNA strängar var identiska till sekvenser från CRISPR lokus. Detta var det första beviset för att CRISPR lokus blir transkriberat till RNA. De undersökte även längden på dessa RNA och fann att vissa av sekvenserna var ibland längre och ibland kortare. De tog detta som ett bevis på att CRISPR lokus transkriberas som ett pre-crRNA (prekursor crRNA) som sedan bearbetas till ett färdigt crRNA. Ytterligare bevis för pre-crRNA hittades av Brouns *et al.* (2008) som såg att ett pre-crRNA innehöll två till tre repetitiva och icke-repetitiva sekvenser från CRISPR. Det vill säga, ett pre-crRNA innehåller runt tre olika spacers när det först transkriberats.



Figur 2. tracrRNA och CRISPR lokus transkriberas. Transkriberat tracrRNA och pre-crRNA binder till varandra och bildar pre-crRNA:tracrRNA duplex. Bilden är omarbetad från Karvelis *et al.* (2013).

En av de viktigaste delarna i bearbetningen av pre-crRNA till crRNA är att det måste binda till tracrRNA (Deltcheva *et al.* 2011). tracrRNA blir transkriberat från en konserverad region som ligger strax framför *cas* generna i genomet (Figur 1.). Det transkriberade tracrRNA:t har visat sig innehålla en sträcka på 25 nukleotider som är identisk med de repetitiva sekvenserna från CRISPR lokus. Man tror att dessa 25 nukleotider basparar med varandra och på så vis bildar ett duplex av tracrRNA och pre-crRNA (Figur 2.) som sedan blir bearbetat av en tredje part (Deltcheva *et al.* 2011). Deltcheva *et al.* (2011) testade teorin om en tredje part genom att mutera tracrRNA och resultaten var att inget pre-crRNA blev bearbetat till crRNA. Resultaten stödjer teorin om en tredje part. I det färdiga duplexet kommer ungefär 22 nukleotider från crRNA:s 3'-ände och 22 nukleotider från tracrRNAs 5'-ände baspara till varandra. Kvar på crRNA finns det 20 nukleotider i 5'-ändan som senare kommer att binda 20 nukleotider komplementärt virus DNA. På tracrRNAs 3'-ände kommer nukleotider bilda veck och interagera med cas9 (Jinek *et al.* 2012).

Den tredje parten som deltar i bearbetningen av pre-crRNA:tracrRNA duplexet tros vara enzymet ribonukleas III (RNase III) som hjälper till i bearbetningen genom att klyva RNA:t på flera ställen. RNase III är inte direkt kopplat till typ II CRISPR/cas systemet, men finns i

bakterierna. RNase III skulle kunna rekryteras inne i bakterien när den väl behövs för att bearbeta pre-crRNA:tracrRNA duplexet till färdigt sgRNA (crRNA:tracrRNA duplex) (Deltcheva *et al.* 2011). För att testa teorin gjorde Deltcheva och hennes team (2011) ett experiment *in vitro*. De skapade pre-crRNA:tracrRNA duplex och blandade det med RNase III. Experimentet gav resultatet att alla ensamma pre-crRNA och tracrRNA inte hade blivit bearbetade till färdigt sgRNA av RNase III, medan duplex hade blivit det. Resultaten kunde stödja teorin om att RNase III bearbetar pre-crRNA och att dessutom interaktionen med tracrRNA är nödvändig för att det ska kunna ske (Deltcheva *et al.* 2011).

#### *sgRNA i cas9*

tracrRNA är förutom i bearbetningen av crRNA även viktigt för igenkänning av den protospacer som ska klyvas. Jinek *et al.* (2012) har utfört experiment på cas9 med och utan tracrRNA för att se hur det påverkade bindningen av protospacers. Resultatet visade en stor skillnad i bindning till cas9 vilket tyder på att tracrRNA har en stor påverkan på hurvida cas9 kan känna igen en protospacer eller inte. En teori är att tracrRNA bidrar med att positionera crRNA rätt så att det kan binda till sin komplementära protospacer (Jinek *et al.* 2012).

En annan viktig del av tracrRNA är dess förmåga att bilda veck i sin 3'-ände. Oftast bildar de mellan ett till tre veck beroende på längden av tracrRNA. Det första vecket verkar vara viktigast för att sgRNA ska kunna binda till cas9, det interagerar med igenkänningsloben och PI domänen av nukleasloben. De andra två vecken binder in till igenkänningsloben och ökar stabiliteten på sgRNA bindningen ytterligare. Den ökade stabiliteten leder även till ökad aktivitet hos cas9 (Nishimasu *et al.* 2014).

## **Bakteriofagers samevolution till CRISPR/cas9**

För att virus ska kunna fortsätta sprida sitt DNA och föröka sig så behöver de ha någon form av samevolution med CRISPR/cas9 för att utveckla ett effektivt försvar. Man har i flera studier funnit att vissa virus har infekterat celler som ska vara resistent mot dem. Deveau med sitt team (2008) undersökte genom hos muterade bakteriofager som infekterat resistent *S. thermophilus*. Detta visade att de bakteriofager som effektivt lyckats infektera *S. thermophilus* antingen hade en eller två utbytta nukleotider i sin protospacer, borttagna nukleotider i sin protospacer eller så hade de en utbytt nukleotid i sin PAM sekvens. Dessa resultat tyder starkt på att bakteriofager undan kommer igenkänningen av PI domänen genom mutationer hos enskilda nukleotider i sitt genom. Andra studier som publicerats senare har dock visat att placeringen av mutationerna har stor betydelse. Om mutationerna sker i 5'-ändan av en protospacer har de inte lika stor påverkan som de har om de sker i 3'-ändan. I 5'-ändan kan upp till sex nukleotider muteras utan att resistensen försvinner (Garneau *et al.* 2010, Jinek *et al.* 2012). En studie av Childs *et al.* (2012) har visat att om mutationer sker i för långsam takt i en bakteriofags protospacer så kan det leda till att den blir utrotad av bakterier som bär komplementära spacers. Det visar på att mutationerna är en mycket viktig del av bakteriofagers gensvar på CRISPR/cas9.

Det är ännu oklart hur CRISPR lokus tar upp protospacers och hur dessa väljs ut, men några studier som visar på att det finns ett mönster i hur det går till. Man har sett att protospacers som väljs ut för integrering i CRISPR lokus ofta angränsar till en sekvens på åtta nukleotider; 5'-GCTGGTGG-3'. Denna sekvens kallas för *Chi* och är mycket vanligt återkommande i bakteriella genom och mindre återkommande i virusgenom (Levy *et al.* 2015). Teorin är att *Chi* gör det lättare för bakterien att skilja på eget DNA och virus DNA. I *Escherichia coli*



finns ett rekombinas som kallas RecBCD som är ett enzymkomplex bestående av både ett helikas och ett nukleas. RecBCD tvinnar upp och degraderar DNA med start från dubbelsträngade DNA brott, och lagar DNA:t genom rekombination. Nukleas aktiviteten håller på fram till det att komplexet stöter på *Chi*, där aktiviteten stoppas och ett enkelsträngat brott induceras på DNA:t där rekombinationen kan påbörja. Att *Chi* är så frekvent i det bakteriella genomet kan vara ett skydd mot att RecBCD inte ska degradera stora delar av det egna genomet hos *E. coli* bakterien (Levy *et al.* 2015).

I studier med *E. coli* har man sett att bakteriofager har utvecklats till att hämma rekombinaset RecBCD eller anta flera *Chi* i sitt genom för att skydda sig mot nukleas aktiviteten hos RecBCD (Bobay *et al.* 2013). *E. coli* har ett CRISPR typ I system vilket kodar för annorlunda *cas* gener är vad CRISPR typ II gör. *E. coli* har alltså inte något cas9, men kodar dock för rekombinaset RecBCD som bidrar i integreringen av nya spacers i dess CRISPR lokus. Genom att undersöka genomet hos bakteriofager såg Bobay med kollegor (2013) att antalet *Chi* är betydligt högre än vad som var förväntat. Eftersom en teori är att mängden *Chi* hjälper bakterier att skilja eget DNA mot virus DNA har man trots att skillnaden i *Chi* mellan bakteriofager och bakterier är ganska stor. Men studier på detta har visat att skillnaden inte alls behöver vara så stor, utan att den beror mycket på vilka egenskaper bakteriofagernas genom innehåller (Bobay *et al.* 2013). Till exempel, om bakteriofager inte kan koda för egna rekombinas behöver de använda sig av rekombinaset som finns i värden de infekterar. I *E. coli* innebär det att bakteriofagernas genom måste innehålla många *Chi* sekvenser för att inte bli degraderat av RecBCD (Bobay *et al.* 2013). Virus som kodar för egna rekombinas brukar även koda för diverse funktioner som kan hämma aktiviteten hos RecBCD. Det finns dock bakteriofager som inte hämmar RecBCD, dessa sitter då i samma situation som de bakteriofager som inte har rekombinas och de behöver ha många *Chi* sekvenser i sitt genom för att det inte ska bli degraderat. Bakteriofager som hämmar RecBCD men inte kodar för egna rekombinas kommer inte att kunna laga de dubbelsträngade brott som uppstår i DNA:t (Bobay *et al.* 2013). Då *Chi* sekvensen är kopplad till just RecBCD bör det bara vara närvaro av RecBCD rekombinaset som avgör om ett genom innehåller stort antal *Chi* sekvenser eller inte. RecBCD har en skyddande roll i *E. coli* genom att degradera bakteriofagers DNA, men hjälper även till i integreringen av spacers. Det är ett rekombinas som finns i *E. coli* och samevolutionen till detta rekombinas har därför bara skett hos de bakteriofager som infekterar *E. coli*, till exempel bakteriofagen lambda. Alla prokaryoter har dock rekombinas och dess virus har sannolikt utvecklats för att antingen använda eller undantrycka dessa rekombinas. Bakteriofagers anpassningar till RecBCD skulle därför kunna appliceras på alla virus i avseende på evolutionen mot rekombinas.

Flera studier har utförts på samevolutionen mellan virus och bakterier. Bland annat en studie från Kashiwagi och Yomo (2011) som starkt visar på att det finns ett samband mellan bakteriers evolution och virus evolution. Studien tar inte upp samevolution med just CRISPR/cas systemet, men visar ändå tydligt på att det finns en samevolution och att samevolutionen riktar in sig på att mutera specifika gener som påverkar bakteriofagers förmåga att infektera och bakteriers förmåga att förhindra infektion. Studien visade också tydligt på att mutationshastigheten hos bakteriofager accelererar när de samevolverar med bakterier, i jämförelse med när bakterier är frånvarande (Kashiwagi & Yomo 2011).

CRISPR/cas9 är ett RNA och DNA baserat immunsystem hos prokaryoter. Eukaryoter har ett liknande, RNA och DNA baserat, immunsystem kallat RNA interferens (RNAi). Det fungerar på liknande vis och vägleder nukleas till komplementärt RNA som blir degraderat (Holoach & Moazed 2015). RNAi kan även skapa en förändring hos kromosomerna vilket gör att

förändringen blir ärftlig, precis som CRISPR lokus hos prokaryoter (Holoch & Moazed 2015). Dessa likheter kan tyda på att RNAi har utvecklats från CRISPR/cas systemet hos prokaryoter. Det fungerar på liknande vis men är på en molekylär och genetisk nivå mycket olika. CRISPR/cas systemet har inte kunnat föras vidare till eukaryoter på grund av att vissa cas proteiner är toxiska, och kan utgöra en fara för eukaryota celler där det är en risk att dessa gener aktiveras på grund av sönderfall av operoner (Koonin 2017). Hos prokaryoter tror man att toxinerna kopplar ihop CRISPR/cas systemet med programmerad celldöd för att agera som en plan B om virus DNA skulle ta sig förbi andra immunsystem (Makarova *et al.* 2012). Koonin (2017) föreslog utefter de studier som finns att dessa två immunsystem hos prokaryoter och eukaryoter har utvecklats oberoende av varandra, trots vissa ytliga likheter.

## Diskussion

Virus och prokaryoter samevolverar med varandra vilket leder till en balans hos bådas populationer. Studier visar att blandningar av bakteriofager och bakteriepopulationer går mot en balans där båda populationerna samlever med varandra (Haerter & Sneppen 2012). CRISPR/cas är ett väl utvecklat adaptivt immunsystem som har evoluerat fram för prokaryoter och dess behov av skydd från virus. Studier av Kashiwagi och Yomo (2011) har visat att mutationshastigheten går fortare när bakteriofager får samevolvera med bakterier. Evolutionen verkar även gå snabbare om den involverar skydd mot infektioner eller andra angrepp (Paterson *et al.* 2010).

Det är en ständig kapplöpning om vem som hinner utvecklas snabbast för att organismen ska få högre fitness. CRISPR/cas systemet är förutom adaptivt även ärftligt vilket gör att bara en cell behöver bli infekterad av ett virus för att sedan alla avkommor av cellen ska ha ett skydd mot denna. Virus har alltid gett ett gensvar på prokaryoternas evolution, i det här fallet att mutera PAM eller närliggande nukleotider för att kunna undkomma CRISPR/cas9:s igenkänningssystem. Något som kan hämma prokaryoters immunförsvar är att det verkar finnas begränsat med plats i CRISPR lokus. Studier på detta har visat att när nya spacers integreras så kommer gamla att tas bort (Deveau *et al.* 2008). Därför finns inte resistensen mot ett virus kvar för alltid. Det finns dock mycket som hämmar virus evolution också. På grund av dess mycket enkla genom kan de behöva göra vissa avvägningar när det kommer till att anta nya egenskaper. Om genomet blir för stort kan det ta längre tid för det att läsas av, och på så vis tar det även längre tid för dem att föröka sig (Bull *et al.* 2004). Då det enda av ett virus som faktiskt kommer in i de prokaryota cellerna är dess DNA så är det mycket viktigt att allt som behövs för att bygga upp nya virus och eventuellt undertrycka prokaryotens immunförsvar finns inlagt i dess genom.

## Lamarckiska evolutionsteorin och CRISPR

CRISPR/cas systemet kan vara ett av få bevis för Lamarckisk evolution. Att CRISPR/cas utvecklas genom att genotypen ändras efter angrepp från virus och att denna genotyp förs vidare till dotterceller stämmer mycket väl överens med den Lamarckiska evolutionsteorin. Enligt denna evolutionsteori har miljön stor påverkan på både genotyp och fenotyp ändringar hos organismer och dessa genotyper förs sedan vidare till avkomman. En moderniserad omformulering av Lamarckiska evolutionsteorin skulle vara att miljöfaktorer inducerar en mutation, mutationen är riktad till en specifik gen och orsakar en förändring som ska förebygga orsaken till att mutationen uppstod i första början (Koonin & Wolf 2009). Det verkar vara just detta som sker med CRISPR/cas systemet.

Det verkar finnas starkast bevis för Lamarckisk evolution i det här fallet, men kan man helt utesluta den Darwinistiska evolutionsteorin? Den teorin säger att evolutionen sker med

slumpmässiga mutationer i genomet och att dessa mutationer kan väljas ut till att föras vidare eller inte genom det naturliga urvalet. Det är inte omöjligt att detta skulle kunna vara fallet även för CRISPR/cas, men de studier som gjorts visar tydligast på att spacers väljs ut snarare än att de muteras fram slumpmässigt. Enligt den Darwinistiska teorin hade alltså en cell kunnat anta en viss spacer utan att först behöva bli infekterad av ett virus, vilket har testats i studier av Haerter och Sneppen (2012). De utförde en studie där de kunde se resultaten av att inte utsätta bakterier för bakteriofager och sedan sekvensera spacers i dess CRISPR lokus. Det visade att integrerandet av spacers går mot extremer, det vill säga att bakterier i en population antingen har full resistens mot en bakteriofag eller ingen resistens alls. Detta visar dock på att bakterierna kan anta spacers även utan att ha dessa i sin omgivning. Man skulle kunna se på dessa studier som ett bevis för att både Lamarcks och Darwins evolutionsteorier samverkar, och att det helt enkelt inte går att helt utesluta varken den ena eller den andra teorin.

Man skulle också kunna se på CRISPR/cas som ett immunsystem likt vår eget. Människor har också ett adaptivt immunförsvar som är ärftligt, vilket skulle kunna jämföras med CRISPR/cas. Människors adaptiva immunförsvar är något som under lång tid har evoluerat fram till vad det är idag, en resa vars resultat förmodligen har formats av både naturligt urval och genomtänkt anpassning till omgivningen. Alltså liknar utvecklingen något av en blandning mellan Lamarcks och Darwins teorier om evolutionen.

### **En avvägning av anpassningar**

Samevolutionen mellan bakteriofager och bakterier med avseende på CRISPR/cas9 ser mycket olika från bakterie till bakterie. De kan alla ha CRISPR/cas9 men sedan också andra typer av immunförsvar som varierar mellan dem. En bakteriofags evolution mot CRISPR/cas9 måste därför även innehålla anpassningar till andra eventuella immunsystem som finns runt om. Den självklaraste utvecklingen man har sett hos bakteriofager gentemot CRISPR/cas9 är mutationerna som sker i deras protospacers och PAM sekvenser som minskar risken för igenkänning av cas9. Detta är en enkel form av utveckling som dessutom inte tar upp någon extra plats i virusens genom, vilket annars är ett problem som virus har när det gäller att samevolvera med sin värd. En avvägning måste ske gentemot vilka anpassningar som är viktigast för överlevnad. En fråga som kvarstår är hur dessa mutationer och anpassningar går till, sker det slumpmässigt eller är det en mer riktad och planerad process? Att veta exakt hur samevolution mellan arter fungerar kan ge oss ett bra verktyg för att kunna förutspå vår egen evolution och utveckling. Hur organismer samevolverar med immunförsvar kan ge oss mer kunskap om utvecklingen av vårt eget immunförsvar. Kanske kan vi i framtiden styra evolutionen av vårt immunförsvar till att ge oss ett bättre skydd från bakterier och virus. I en tid som denna när vår överanvändning av antibiotika har orsakat att vi inte längre kan bli botade från alla infektioner är det en upptäckt som kan vara mycket viktig för vår överlevnad.

### **Tack**

Tack till Sophia Forsskåhl och Stefan Public som har bidragit med kommentarer och synpunkter.

### **Referenser**

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

- Bobay LM, Touchon M, Rocha EPC. 2013. Manipulating or Superseding Host Recombination Functions: A Dilemma That Shapes Phage Evolvability. *PLoS Genetics*: doi: 10.1371/journal.pgen.1003825
- Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.
- Bull JJ, Badgett MR, Springman R, Molineux IJ. 2004. Genome properties and the limits of adaptation in bacteriophages. *Evolution* **58**: 692-701.
- Childs LM, Held NL, Young MJ, Whitaker RJ, Weitz JS. 2012. Multiscale model of CRISPR-induced coevolution dynamics: diversification at the interface of Lamarck and Darwin. *Evolution* **66**: 2015–2029.
- Chylinski K, Rhun AL, Charpentier E. 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA biology* **10**: 726–737.
- Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602–607.
- Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008. Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *Journal of bacteriology* **190**: 1390-1400.
- Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.
- Haerter JO, Sneppen K. 2012. Spatial Structure and Lamarckian Adaptation Explain Extreme Genetic Diversity at CRISPR Locus. doi: 10.1128/mBio.00126-12
- Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA. 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**: 199-202.
- Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. *Nature reviews genetics* **16**: 71–84.
- Hou Z, Zhang Y, Propson NE, Howden SE, Chu L, Sontheimer EJ, Thomson JA. 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proceedings of the national academy of science of the United States of America* **110**: 15644–15649.
- Jansen R, van Embden JDA, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **43**: 1565-1575.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**: 816-821.
- Karvelis T, Gasiunas G, Miksys A, Barrangou R, Horvath P, Siksnys V. 2013. crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA biology* **10**: 841-851.
- Kashiwagi A, Yomo T. 2011. Ongoing Phenotypic and Genomic Changes in Experimental Coevolution of RNA Bacteriophage Q $\beta$  and *Escherichia coli*. *PLoS genetics*: doi: 10.1371/journal.pgen.1002188
- Koonin EV. 2017. Evolution of RNA- and DNA-guided antiviral defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. *Biology direct* **12**: 5.
- Koonin EV, Wolf YI. 2009. Is evolution Darwinian or/and Lamarckian?. *Biology Direct* **4**: 42.
- Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, Edgar R, Qimron U, Sorek R. 2015. CRISPR adaptation biases explain preference for acquisition of foreign DNA.

- Nature **520**: 505–510.
- Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata S, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**: 935–949.
- Makarova KS, Anantharaman V, Aravind L, Koonin EV. 2012. Live virus-free or die: coupling of antiviral immunity and programmed suicide or dormancy in prokaryotes. *Biology direct* **7**: 40.
- Manica A, Zebec Z, Teichmann D, Schleper C. 2011. In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Molecular microbiology* **80**: 481–491.
- Mojica FJM, Díez-Villaseñor C, Soria E, Juez G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology* **36**: 244–246.
- Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. 2005. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular evolution* **60**: 174–182.
- Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733–740.
- Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libberton B, Fenton A, Hall N, Brockhurst MA. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature* **464**: 275–278.
- Tang TH, Bachellerie JP, Rozhdetsvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Hüttenhofer A. 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proceedings of the national academy of science of the United States of America* **99**: 7536–7541.

# CRISPR/cas9: funktion och bakteriofagernas samevolution: etisk bilaga

**Carolina B. Viman**

Självständigt arbete i biologi 2017

## **Att tillåta genmodifiering av könsceller**

På senaste tiden har mycket hänt inom utvecklingen av olika metoder för genmodifiering, till exempel upptäckten och användandet av CRISPR/Cas9. Det ger oss större chanser att modifiera bort till exempel genetiska sjukdomar. Men ännu är det inte lagligt att genomföra någon typ av genmodifiering i könsceller för att få bort genetiska sjukdomar hos foster. Bör vi använda CRISPR/Cas9 på könsceller för att få bort dessa genetiska sjukdomar?

För många föräldrar som är bärare av allvarliga genetiska sjukdomar kan det vara ett jobbigt val om att skaffa barn eller inte då risken finns att barnet utvecklar denna sjukdom som de bär på. Med genmodifiering med CRISPR/Cas9 skulle man kunna ta bort den genen som kodar för den genetiska sjukdomen och på så sätt öka föräldrarnas chans att få ett friskt barn. Att tillåta denna typ av genmodifiering skulle också vara positivt ut en ekonomisk synpunkt. Individer som spenderar hela sina liv in och ut ur sjukhus är en stor ekonomisk belastning för samhället. Utan denna ekonomiska belastning skulle mera pengar kunna läggas på medicinsk forskning som kan hjälpa att bota andra icke genetiska sjukdomar.

Negativa effekter av att tillåta genmodifiering av könsceller är att genmodifieringen då förs vidare till flera generationer. Framtida generationer kan inte ge ett samtycke till detta, och vi kan heller inte ta oss friheten till att genomföra genmodifiering utan samtycke från de som blir påverkade av det. Genmodifieringen kan även ha långsiktiga effekter som kan vara svåra att förutse. Det är även svårt att dra gränsen på vilka sjukdomar som ska bort och inte och att avgöra vad som ska klassas som en allvarlig sjukdom, och vem som ska ta det beslutet. Att låta sjukvården dra gränsen istället för föräldrarna kan minska risken för att genmodifiering börjar användas till annat än mot sjukdomar, då ett vanligt motargument mot genmodifiering är att det kan börja användas till att ändra mindre viktigare egenskaper som utseende eller fysik. Samtidigt har vi inte kunskaper nog att garantera att genmodifieringen inte kommer att leda till andra bieffekter, risken finns därför att vi modifierar bort en sjukdom men istället skapar en annan.

Det finns många motargument mot användningen CRISPR/cas9 i könsceller som fortfarande går obesvarade. CRISPR/Cas9 är ännu ett ganska nytt verktyg inom genmodifiering, och mera forskning kommer definitivt att behövas innan vi tar kan ta steget. Därför bör vi inte använda CRISPR/cas9 på könsceller.