

Slutrapport (juni 2016)

MHC – Centrum av Hästens Immunsystem

(Projektnummer H1147283)

Tomas Bergström

Sofia Mikko Göran Andersson

BAKGRUND

Major histocompatibility complex (MHC) är en region i hästens genom som utgör ett 4 miljoner baspar (Mbp) långt område på kromosom 20 och innehåller ett stort antal gener med viktiga funktioner för immunsystemet. I flera studier av ärftliga sjukdomar hos häst har en koppling till MHC regionen påvisats. Till skillnad från människan och musen, är denna region av hästens arvsmassa dåligt karakteriserad både vad gäller antal gener och hur stor genetisk variation som finns i olika loci. Det övergripande målet med detta projekt har därför varit att karakterisera hästens MHC region för att därigenom ge bättre förutsättning för studier av komplexa immunologiska sjukdomar.

I den ansökan som beviljades sökte vi bland annat anslag för en 50 % finansiering av en laboratorieassistent. Tack vare medfinansiering av fakulteten för veterinärmedicin och husdjursvetenskap (VH) kunde en doktorand antas under 2013, för fyra års forskarutbildning. Nedan följer en slutrapport av det arbete som har bedrivits med stöd av det beviljade anslaget från Stiftelsen hästforskning samt vad som planeras under doktorandens fjärde år av forskarutbildning. Disputationen beräknas att ske under mitten av 2017.

MATERIAL OCH METODER

Under 2013 skedde en insamling av blodprover från hästar som sedan genotypades med mikrosatelliter för att identifiera individer med relativt hög grad av homozygoti över MHC-regionen. Detta för att minska problematiken med att samtidigt sekvensera två olika haplotyper. Två hästar valdes ut för helgenomsekvensering med "next generation sequencing (NGS) teknologi. De två hästar som sekvenserades var ett svenskt varmblod (SWB) och en färöisk ponny. Sekvenseringen gjordes på NGS-plattform IonProton med målsättning att erhålla 6-8 X täckning av genomet per individ.

Under 2014 fokuserades arbetet dels på att sekvensera fler hästraser och dels på att förbättra referensgenomet för häst när det gäller MHC-regionen. Inledningsvis genomfördes helgenomsekvensering av Akhal-Teke hästar med två NGS-tekniker. Helgenomsekvenseringen genomfördes dels på Illumina-plattformen som i likhet med IonProton ger korta sekvenser (2x150 bp) och dels på SOLID-plattformen. Den

senare ger också korta sekvenser (reads) men sekvensbiblioteken är flera kilobaser (kbp) och på så vis finns möjlighet att också detektera duplikationer i genomet.

Ett problem med MHC regionen är att flera områden i MHC-regionen är duplicerade. Referensgenomet som bygger på sekvensering av fullblodshästen Twilight och finns i de publika databaserna är därför inte helt tillförlitligt i MHC-regionen. Under 2014 inledde vi därför ett sekvenseringsprojekt av BAC-kloner (bacterial artificial chromosomes) som täcker MHC-regionen. BAC-klonerna härrör från hästen Bravo, ett halvsyskon till hästen Twilight som är den häst vars DNA är det befintliga referensgenomet för häst. Vår samarbetspartner prof. Loren Skow vid Texas A&M har i en studie gått igenom samtliga BAC-kloner från Bravo och valt ut 31 BAC-kloner som sannolikt täcker hela MHC-regionen. Dessa 31 BAC-kloner översändes till oss och bakteriekulturer med BAC-kloner sattes upp för att erhålla tillräckliga mängder för sekvensering.

DNA från BAC-klonerna renades fram från åtta av dessa för sekvensering med "long-read single molecule, real-time (SMRT) sequencing technology" från företaget Pacific Biosciences (PacBio). Det finns två fördelar med detta tillvägagångssätt. Dels vet vi från vilken BAC-klon sekvenserna härrör från och dels ger PacBio till skillnad från andra NGS-plattformar långa sekvenser (long reads). Istället för "reads" med en längd av ca 150 baspar (bp) sekvens ger PacBio "reads" som är flera tusen baspar (kb).

Under 2015 och 2016 har ett intensivt arbete genomförts för att producera en tillförlitlig sekvens av hästen MHC Class II region. Sekvensdata från PacBio av åtta BAC-kloner kompletterades med long-range PCR och PacBio sekvensering för att kunna producera en sammanhängande sekvens över hela Class II regionen. Därefter vidtogs omfattande bioinformatiskt annoteringsarbete för att producera en ny referenssekvens med säker annotering av både proteinkodande gener och pseudogener i regionen. För annoteringspipelinen användes RNA-seq data från olika vävnadstyper samt bioinformatiska programvaror för genannotering.

RESULTAT

Resultaten från sekvenseringen på IonProton plattformen gav en genomsnittlig

täckning av genomet på ca 10X. Det innebar att varje häst blev sekvenserad så att varje nukleotid i genomet i genomsnitt sekvenserades ca 10 ggr. Totalt genererades mellan 180-200 miljoner "reads" per häst och varje "read" var i medeltal 125 bp lång.

Sekvenserna från de två hästarna jämfördes sedan med referenssekvensen (EquCab 2.73). Analysen visade att en täckning av genomet (ca 10X) var tillräckligt hög för att fånga upp genetiska variationen hos de undersökta hästarna. En stor del av MHC-regionen stämde väl överens mellan referenssekvensen från fullblodet Twilight och de två hästar som vi sekvenserade. Analysen tydde dock på att det finns flera områden i hästens MHC där det svenska varmblodet och framförallt den färöiska ponnyn skiljer sig åt. Bland annat uppvisade MHC-DRB regionen skillnader som kan tyda på att antalet DRB-gener varierar mellan individer. Det blev därmed klarlagt att sekvensering med korta "reads" inte är tillräckligt fungerar för att detektera kopietalsvariation i MHC regionen.

Under 2014 inriktades därför arbetet på förbättra referensgenomet i MHC-regionen genom att bland annat sekvensera BAC-kloner från hästen Bravo. Inledningsvis sekvenserades fyra BAC-kloner på PacBio systemet. Resultaten visade att tekniken producerade tillräckligt långa sekvenser av hög kvalitet för att generera en tillförlitlig referenssekvens. Ytterligare fyra BAC-kloner sekvenserades därför med samma teknik. Resultaten visade att två av BAC-klonerna inte överlappade med varandra. För att kunna producera en sammanhängande sekvens över hela hästens Class II region utan mellanliggande okänd sekvens mellan de två BAC-klonerna användes "long-range PCR" för amplifiering av området. På så vis kunde regionen mellan de två BAC-klonerna överbryggas. PCR-produkterna sekvenserades på PacBio-plattformen och en sammanhängande högkvalitativ sekvens över hela hästens Class II region kunde konstrueras. Totalt sett spände hela sekvensen över 1.2 miljoner baspar (Mbp).

Genannoteringen av hästens MHC Class II region resulterade i en identifiering av fler än 30 gener samt en noggrann kartläggning av vilka av dessa som kunde klassificeras som proteinkodande gener och pseudogener. En viktig del av resultatet från genannoteringen var att definiera genomiska koordinater för transkriptionsstart, exon och intron. Dessutom genomfördes en annotering av andra genomiska enheter (features) så som repetitiva element. En komparativ analys av hästens och människans MHC Class II region visade att hästens genstruktur skiljer sig radikalt från

människans. Ett manuskript som beskriver resultaten kommer inom kort att skickas in till en vetenskaplig tidskrift granskning.

Arbete kommande året

Under det fjärde året av doktorandens forskarutbildningen kommer arbetet att koncentreras kring studier av haplotypvariationen och graden av polymorfi i hästens MHC Class II region. Studier från andra arter har visat att bland annat antalet MHC-DRB gener varierar på olika kromosomer och därmed skiljer sig åt mellan individer. Baserat på den nya referenssekvensen har vi nu inlett detta arbete. Vi avser även att utveckla ett genotypningssystem för hästens MHC.

DISKUSSION

Gener i MHC-regionen spelar en central roll i immunsystemet och det är väl känt att regionen är associerad till ett stort antal immunomedierade sjukdomar. Det övergripande målet med detta projekt har därför varit att karakterisera hästens MHC region och därigenom ge förbättrade förutsättning för studier av inflammatoriska, infektiösa och autoimmuna sjukdomar. Den nya referenssekvensen av hästens MHC Class II region är baserad på avancerad sekvenserings teknologi (long-read single molecule, real-time (SMRT) sequencing technology) av BAC-kloner på en plattform utvecklad av företaget Pacific Biosciences (PacBio). Detta resulterade i en högkvalitativ och heltäckande sekvens som spänner över hela hästens MHC Class II region (1.2 Mbp). Den påföljande genannoteringen ger för första gången en gedigen genetisk karta med exakta koordinater av regionens dryga 30 gener. Det är en nödvändig uppdatering av en av de viktigaste regionerna i hästens arvs massa och resultaten bidrar därför med avgörande information för studier av immunomedierade sjukdomar hos hästen. Projektet hade inte varit möjligt att genomföra utan tillgång till sekvensering av BAC-kloner med långa ”reads”. Våra initiala studier som baserades på annan sekvens teknologi med korta ”reads” (IonProton) visade att det var möjligt att fånga upp variation i hästens MHC men däremot omöjligt att definiera till vilken gen den genetiska variationen är kopplad till. Det berodde sannolikt på att variationen i hästens MHC är relativt hög, både vad gäller antalet paraloga gener och antalet alleler vid flera loci. Med en ny och

tillförlitlig referenssekvens som grund är det nu möjligt att studera hästens MHC-variation och använda den kunskapen för studier av inflammatoriska, infektiösa och autoimmuna sjukdomar. Av den anledningen fokuserar vi nu på att definiera haplotyp-variationen, graden av polymorfi och att utveckla ett genotypningssystem för hästens MHC.

PUBLIKATIONER

Nedan redogörs för projektets publikationer och presentationer i vid konferenser och i vetenskapliga tidskrifter.

Abstrakt och vetenskapliga presentationer

Den första helgenomsekvensering som baserades på korta ”reads” som genererades med IonProton plattformen presenterades på ”The 10th Dorothy Russell Havemeyer Foundation International Equine Genome Mapping Workshop” som hölls mellan den 10-13 juli, 2013 på Azorerna.

Preliminära resultat från PacBio sekvenseringen av BAC-kloner presenterades vid ”The International Society of Animal Genetics (ISAG 2014)”, i Xián, Kina den 28 juli till 1 augusti

En första version av den nya och annoterade referenssekvensen för hästens MHC Class II region presenterades under “The 11th Dorothy Russell Havemeyer Foundation International Equine Genome Mapping Workshop” den 22-25 juli, 2015 i Hannover, Tyskland.

Vetenskapliga tidskrifter

Viluma A. et al. *The equine major histocompatibility complex class II region resolved using PacBio long-read sequencing technology.* (2016 under färdigställande).

SLUTSATSER

Hästens MHC region skiljer sig radikalt från människan både vad avser antal gener i Class II regionen och ordningen på dessa gener. Den noggranna annoteringen av hästens MHC och en högkvalitativ referenssekvens är nödvändig för studier av komplexa immunologiska sjukdomar hos häst. Bland de sjukdomar som har en koppling till MHC kan nämnas sarcoider (inkar), sommareksem, allergier, infektionssjukdomar samt vissa former av cancer med. Under det närmaste året kommer vi att fokusera på att definiera variation i hästens MHC samt undersöka möjligheten att utveckla ett genotypingssystem för hästens MHC.

RESULTATFÖRMEDLING

Resultaten har förmedlats vid olika vetenskapliga konferenser så som ”Dorothy Russell Havemeyer Foundation International Equine Genome Mapping Workshop” och “The International Society of Animal Genetics (ISAG). När resultaten är publicerade i vetenskapliga tidskrifter avser vi även att skriva populärvetenskapliga artiklar och att sprida informationen i ett bredare sammanhang. På kurser och program vi vid SLU använder resultaten och erfarenheten av detta projekt i undervisningen.