

## SAMENVATTING

### HET KRAKEN VAN GENOMEN VAN BODENARTHROPODEN.

Evertebrate bodemgemeenschappen zijn opvallend divers en essentieel voor de gezondheid van het bodemecosysteem. Collembola (ook wel springstaarten genoemd) zijn vleugellose zespotige ongewervelde dieren die bodemmineralisatie ondersteunen, doordat ze zich voeden met dood organisch materiaal (detritus) dat vervolgens beter beschikbaar wordt voor mineraliserende microbiële gemeenschappen in de grond. Hoewel ze intensief worden bestudeerd in de context van ecotoxicologie en bodemecologie, is er vrijwel niets bekend over de genomische samenstelling van deze organismen. Dit proefschrift presenteert genomisch onderzoek aan twee modelspringstaarten: *Folsomia candida* en *Orchesella cincta*.

De doelstellingen van dit onderzoek waren 1) het assembleren en annoteren van de transcriptomen en genomen van *F. candida* en *O. cincta*. 2) karakterisatie van de genomische inhoud, 3) meer inzicht verkrijgen in de evolutie van springstaart-genomen, en 4) het beter begrijpen van specifieke kenmerken en mogelijke aanpassingen en pre-adaptaties gerelateerd aan het bodemleven en omgevingsstress.

Hoofdstuk 2 beschrijft assemblage en annotatie van van *F. candida* en *O. cincta* transcriptomen. Tevens wordt een vergelijkende analyse gepresenteerd van eiwitcoderende gensequenties in springstaarten en hun orthologen in kreeftachtigen (Crustacea) en insecten. Op die manier kunnen we achterhalen welke eiwitten in verband kunnen worden gebracht met adaptieve evolutie van hexapoden binnen de clade van de Pancrustacea. Een statistische test om positieve selectie in Hexapoda te detecteren werd op alle orthologenclusters uitgevoerd. Het was ook mogelijk om eiwitten te identificeren die specifiek onder positieve selectie staan in springstaarten. Significante positieve selectie in Hexapoda werd gevonden voor 250 eiwitten en 57 eiwitten vertoonden positieve selectie in Collembola. Als we de biologische functies bekijken van deze eiwitten dan blijken deze betrokken bij onder andere stofwisseling, stress respons (i.e. DNA-reparatie, immuunrespons), ionentransport, ATP metabolisme, regulering en ontwikkelingsbiologische processen (i.e. oogontwikkeling, neurologische ontwikkeling). De conclusie is dat de geïdentificeerde genfamilies processen vertegenwoordigen die een belangrijke rol gespeeld hebben tijdens de afsplitsing van hexapoden vanuit de kreeftachtigen. Deze afsplitsing heeft uiteindelijk geleid tot de meest soortenrijke groep van alle dieren: de Hexapoda. De bovengenoemde processen die geassocieerd zijn met de evolutie van de Hexapoda geven bovendien meer inzicht in de manier waarop deze groep vanuit een marine leefomgeving op het land is geëvolueerd.

In hoofdstuk 3 presenteren wij de eerste genomsequentie van een *O. cincta* laboratoriumpopulatie, die een geschatte omvang heeft van 286,8 miljoen baseparen (Mbp). Hierin zijn 20.249 eiwit-coderende genen ontdekt. Veel van die genen zijn gerangschikt in genfamilies, waarbij 446 van deze families groter zijn geworden tijdens de evolutie van *O. cincta*. Ook identificeerden wij 1169 genfamilies die alleen in deze soort zijn ontstaan. Veel van deze genfamilies zijn betrokken bij algemene biologische processen, zoals transcriptie, translatie en energiemetabolisme. Maar we vonden ook genclusters die betrokken zijn bij xenobiotische biotransformatie. Bovendien identificeerden wij 253 gevallen van horizontale genoverdracht (HGO), een proces waarbij genetisch materiaal wordt overgedragen tussen ongerelateerde organismen. Hoewel het grootste deel van de HGO-genen afkomstig zijn uit bacteriën (37,5%), vonden wij een onverwacht hoog percentage (30,4%) van dergelijke genen afkomstig van schimmels. Opvallend veel HGO-genen zijn betrokken bij

koolhydraatmetabolisme en degradatie van cellulose. Bovendien zijn sommige HGO-genen (bijv. bacillopeptidasen) geëvolueerd als genfamilie na horizontale genoverdracht, waardoor ze sterk in aantal zijn toegenomen in *O. cincta*. Onze hypothese is dat horizontaal overgebrachte genen voordelig zijn voor voedselvertering in het bodemecosysteem dat vol zit met rottend organisch materiaal. Tenslotte ontdekten wij diverse soortspecifieke genen, geëxpandeerde genfamilies en HGO genen, die geassocieerd zijn met differentiële genexpressie als gevolg van genetische aanpassing aan metaalverontreiniging in de grond. Onze hypothese is dat deze genoomfuncties pre-adaptaties kunnen zijn, waar natuurlijke selectie op kan acteren. Ten slotte biedt deze genoomstudie een solide basis voor verdere analyse van evolutionaire mechanismen van aanpassing aan milieustressoren.

In hoofdstuk 4 beschrijven wij het eerste referentie-genoom voor *F. candida*, dat 221,7 Mbp groot is en 28.734 genen bevat. De genomsequentie heeft een zeer hoge kwaliteit en bestaat uit slechts 162 “scaffolds” (de grootste ononderbroken DNA-sequenties die door de assemblageprocedure samengesteld kunnen worden), waarbij het langste scaffold 28,5 Mbp telt. Om specifieke genomische functies gekoppeld aan het bodemleven te onthullen, bestudeerden wij expansie van genfamilies-, soortspecifieke genfamilies, en genen verworven door HGO. Het unieke bouwplan van dit dier inspireerden ons om ook het *Hox*-gecomplex te bestuderen om de typische morfologie van *F. candida* te verklaren, die gevormd is door aanpassingen aan het bodemleven. We vonden 368 genfamilie-expansies, 74 soortspecifieke genclusters, evenals 10.080 genen waarvoor we geen homologie vinden in enig ander organisme. *Folsomia*-specifieke genfamilies zijn geassocieerd met stress-responsen. Verder zijn er 809 (2,8%) gevallen van HGO geïdentificeerd. De belangrijkste bronnen van HGO zijn bacterieel (39,9%), gevolgd door schimmels en protisten. De meeste HGOs zijn betrokken bij koolhydraatmetabolisme, een patroon dat eerder in *O. cincta* is waargenomen.

*F. candida* bevat 11 *Hox* genen binnen één cluster. Echter, door translocatie is de volgorde van de *Hox*-genen ten opzichte van elkaar sterk veranderd ten opzichte van de volgorde die we in andere geleedpotigen vinden. Bovendien identificeerden wij uitgebreide gebieden van colineariteit (gekopieerde clusters met genen die op andere plekken in het genoom weer zijn ingebouwd en herkenbaar zijn aan een overeenkomstige genvolgorde) tussen een aantal genomische clusters van *F. candida*. Tenslotte, was het mogelijk om 'de volledige genomsequentie van *Folsomia*'s endosymbionte bacterie *Wolbachia* te assembleren en ontrafelen. Dit bacteriële genoom blijkt aanzienlijk groter te zijn dan enig ander onderzocht *Wolbachia*-genoom tot nu toe.

De uitbreiding van genfamilies gerelateerd aan stressresponsen suggereert dat aanpassing aan stress een belangrijk kenmerk is voor kolonisatie van de bodem. Ook het relatief grote aantal HGO-genen betrokken bij koolhydraatmetabolisme, in het bijzonder de afbraak van lignocellulose, is waarschijnlijk geëvolueerd in *Folsomia* als aanpassing aan bodemleven. De bodem zit namelijk vol met afbraakproducten van celwanden van planten, schimmels en bacteriën. De colineariteitanalyse van het *F. candida* genoom suggereert dat de sommige genoomstructuren zijn ontstaan als gevolg van ongeslachtelijke (parthenogenetische) voortplanting die we in *Folsomia* waarnemen. Parthenogenese is een manier van voortplanting die veel voorkomt bij geleedpotigen in de bodem. Bij *Folsomia* is de parthenogenese waarschijnlijk opgelegd door *Wolbachia*. Zoals eerder vermeld is *Folsomia*'s endosymbiotische *Wolbachia*-genoom het grootste dat tot nu toe is geïdentificeerd. Dat is opmerkelijk, omdat vaak wordt waargenomen dat deze genomen sterk reduceren als gevolg van horizontale genoverdracht naar het gastheergenoom. In ieder geval levert dit referentiegenoom van *F.*

*candida* een essentiële bron voor verder wetenschappelijk onderzoek naar bijvoorbeeld aanpassingen aan het bodemleven.

Er kan geconcludeerd worden dat de data gepresenteerd in dit proefschrift een solide basis vormt voor verdere vergelijkende genomische studies van springstaarten. Tevens zal het nieuwe inzicht geven in de evolutie van Collembola. Deze studie biedt ook een basis voor verdere analyse van evolutionaire mechanismen van aanpassing aan milieustressoren en het leven in de bodem. Tot slot hebben wij voor beide genomen een genoombrowser ontwikkeld en hebben wij alle genoom- en transcriptoominformatie beschikbaar gesteld voor andere onderzoekers (zie de website [www.collembolomics.nl](http://www.collembolomics.nl)).

