Evolution of Insect Germline Specification

BY

HONGHU QUAN
B.S., Huazhong Agricultural University, Wuhan, China, 2011

THESIS

Submitted as partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Sciences in the Graduate College of the University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:

Jeremy A. Lynch, Advisor
Teresa V. Orenic, Chair
Peter Okkema
Ron Dubreuil
Urs Schmidt-Ott, University of Chicago
ACKNOWLEDGEMENTS

I would first like to acknowledge and thank my thesis advisor, Dr. Jeremy A. Lynch, a wonderful, patient and generous person, an intelligent, knowledgeable and visionary scientist, who I always look up to. It would be impossible to finish my thesis without him. I am really grateful to have him as my mentor to train me as a future independent scientist. Jeremy always gave me a lot of space to think broadly, to do research that I was interested in, and to take time recovering from the failures I had in my projects. He shared many brilliant ideas, thoughts, and interesting papers with me. He provided me many opportunities to attend meetings and workshops. He was very patient and generous to me, because I knew I was not the easiest person to deal with. He used his actions instead of words to show me how a scientist with passion enjoys the research. I will always be grateful for his mentorship and cherish the six years of good memories.

I would also like to thank my committee members, Teresa, Pete, Ron and Urs. Teresa is warmhearted, thoughtful and helpful. As my committee chair, she put my mind at ease when I was nervous at my committee meetings and my preliminary exam. She was always open to talk to me about my research problems. Pete, Ron and Urs, they always gave me very good suggestions to help me make progress on my projects. They asked wonderful questions to broaden my thoughts and make me think critically. Especially Urs, he had to come all the way to UIC from the University of Chicago for my committee meetings. I really appreciated that.
I also want to acknowledge the staff in the Department of Biological Sciences, Beth, Suzanne, Margaret, Judith, Omar, Jacquelyn and Thomas. Thank you for making my work much easier at UIC.

My family shapes who I am. I would like to thank my parents, Jianwen Quan and Tongqin Chang, for teaching me to dream big, encouraging me to see the world, and giving me their unconditional love. They did not like to bother me with their issues, knowing that I would worry about them if they told me. I want to specially acknowledge my mother for her deep love and warm support. I always feel guilty that I could not be with her and take care of her when she was battling diabetes and lost her visions. Also, I want to thank my sister, Yanting Quan, for being such a wonderful sibling. We had a lot of good memories growing up together. Even though she is married now and has a son, she still tries her best to take care of our parents. At last, I want to thank my cousin, Hongyan Quan, who is a good older sister, an independent woman, a successful lawyer and a role model for me. I will always cherish our special bond.

I would like to acknowledge my friends, they are like family to me since I came to U.S. alone and I have no relatives here. They helped me go through the toughest time in my life. They provided unconditional support for me. They listened to my problems and comforted me. I know I can always count on them. First, I want to give special thanks to my long-time friend, Yi Zheng. We knew each other since 2008 in college. We had gone though some hard times studying for GRE and TOFEL. We helped each other prepare for the tests, and we came to U.S. together to pursue our dreams. He is the one who can
always remind me of why I came to U.S. Even though we were not in the same city in U.S., over the years till now we still contact each other every week to share joys, problems and thoughts. Second, I thank my another six years friend, Zhefu Chen, who was my first roommate in U.S. and one of my best friends now. We kept each other company so we did not feel lonely, and we still are. Third, I also want to acknowledge my other old friends back in China, I knew many of them since middle school and high school. They taught me a lot when I was just an immature boy, they always took my side and backed me up, and they believed in me that I can achieve more. We did not spend a lot of time together after I went to college, but we still contact each other once in a while. We had a lot of fun every time I went back home. Last but not least, I want to thank my American friends, who are my classmates, my lab members and my gym buddies. The experience of studying abroad is not only about the degree, it is also about the different cultures. D’Feau, Avinash, we knew each other one month after I came to U.S. They taught me to speak proper English and took me on a road trip to the Great Smoky Mountains. Daniel Pers, who is like my brother, taught me a lot of American slangs and explained a lot of cultural differences for me. We argued a lot, but we also shared a lot of good memories. Zana and Sarah, with whom I can share my life, are always nice and warm to me. Taylor, John and Ali, they are my crossfit coaches and gym buddies. Workout with them definitely helped me relax and take my mind off the work in the lab.
CONTRIBUTIONS OF AUTHORS

Part of Chapter 1 is from the published review article for which I co-authored with my thesis advisor Jeremy A. Lynch (Quan H, Lynch JA: The evolution of insect germline specification strategies. *Current opinion in insect science* 2016, 13:99-105.). Chapter 2 represents a preprint deposited in BioRiv (Quan H, Lynch JA: Transcriptomic and functional analysis of the oosome, a unique form of germplasm in the wasp Nasonia vitripennis. *bioRxiv* 384032; doi: https://doi.org/10.1101/384032.) and submitted to BMC Biology. Jeremy and I collaborated on experimental design, analyzed the data and wrote the paper. I performed the experiments and collected the data. Chapter 3 represents a manuscript in preparation, in which Jeremy and I designed the experiments and analyzed the data, and I performed the experiments and collected the data. Chapter 4 contains the unpublished data, for which Jeremy and I designed the experiments and analyzed the data, and I performed the experiments and collected the data. Chapter 5 summarized my thesis work described in the previous chapters. The Appendix presents the licenses and permissions to re-use the published figures in Chapter 1, Chapter 3 and Chapter 4.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The process of germline development</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Evolution of oskar and the origins of maternal provision in insects</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Germ plasm and maternal provision mode</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 The molecular mechanism of germ plasm assembly in Drosophila</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 The mechanism of germ plasm assembly in Nasonia</td>
<td>7</td>
</tr>
<tr>
<td>1.2.4 The origin of oskar gene corresponds with the appearance of the maternal provision mode in insects</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Mechanisms of zygotic induction</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Evolutionary liability of germline determination modes</td>
<td>15</td>
</tr>
<tr>
<td>1.5 The diversity of germline specification</td>
<td>19</td>
</tr>
<tr>
<td>1.5.1 Germline specification in Drosophila melanogaster and Nasonia vitripennis</td>
<td>19</td>
</tr>
<tr>
<td>1.5.2 Germline specification in Caenorhabditis elegans and Danio rerio</td>
<td>21</td>
</tr>
<tr>
<td>1.6 The evolutionary necessity of the chosen insects</td>
<td>21</td>
</tr>
<tr>
<td>1.7 Thesis goals and organization</td>
<td>24</td>
</tr>
<tr>
<td>1.8 Literature Cited</td>
<td>25</td>
</tr>
<tr>
<td>2. TRANSCRIPTOMIC AND FUNCTIONAL ANALYSIS OF THE OOSOME, A UNIQUE FORM OF GERM PLASM IN THE WASP NASONIA VITRIPPENIS</td>
<td>39</td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>40</td>
</tr>
<tr>
<td>2.1.1 Background</td>
<td>40</td>
</tr>
<tr>
<td>2.1.2 Results</td>
<td>40</td>
</tr>
<tr>
<td>2.1.3 Conclusions</td>
<td>40</td>
</tr>
<tr>
<td>2.2 Background</td>
<td>41</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>46</td>
</tr>
<tr>
<td>2.3.1 RNA-seq analyses of the anterior and posterior poles of the wasp Nasonia early embryos</td>
<td>46</td>
</tr>
<tr>
<td>2.3.2 Novel transcripts localized in the anterior half of the Nasonia early embryos</td>
<td>47</td>
</tr>
<tr>
<td>2.3.4 General description of the novel transcripts localized in the posterior pole of the Nasonia early embryos</td>
<td>49</td>
</tr>
<tr>
<td>2.3.5 Transcripts enriched in the oosome but excluded from pole cells</td>
<td>55</td>
</tr>
<tr>
<td>2.3.6 Transcripts enriched in the posterior pole but not specifically the oosome</td>
<td>58</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>2.3.7 Functional analysis by parental RNA interference showed low phenotypic penetrance</td>
<td>59</td>
</tr>
<tr>
<td>2.3.8 Development of an embryonic injection protocol for Nasonia RNAi</td>
<td>62</td>
</tr>
<tr>
<td>2.3.9 RNAi against three novel germ plasm components unexpectedly disrupts the oosome at from an early stage</td>
<td>64</td>
</tr>
<tr>
<td>2.3.10 RNAi against Nv-coronin and Nv- innexin1 does not affect the oosome, but disrupts pole cell formation</td>
<td>67</td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>72</td>
</tr>
<tr>
<td>2.4.1 RNA-seq analyses</td>
<td>72</td>
</tr>
<tr>
<td>2.4.2 Comparison of the oosome of Nasonia to the polar granules of Drosophila</td>
<td>75</td>
</tr>
<tr>
<td>2.4.3 Unexpected functions of novel oosome components</td>
<td>79</td>
</tr>
<tr>
<td>2.5 Conclusion</td>
<td>84</td>
</tr>
<tr>
<td>2.6 Methods</td>
<td>84</td>
</tr>
<tr>
<td>2.6.1 Sample preparation</td>
<td>84</td>
</tr>
<tr>
<td>2.6.2 RNA sequencing data analysis</td>
<td>86</td>
</tr>
<tr>
<td>2.6.2 Embryonic RNA interference (eRNAi)</td>
<td>87</td>
</tr>
<tr>
<td>2.7 Supplementary materials</td>
<td>88</td>
</tr>
<tr>
<td>2.7.1 Significantly enriched transcripts from both experiments</td>
<td>89</td>
</tr>
<tr>
<td>2.7.2 Command jobs</td>
<td>97</td>
</tr>
<tr>
<td>2.7.3 Primer list</td>
<td>103</td>
</tr>
<tr>
<td>2.8 Literature Cited</td>
<td>107</td>
</tr>
<tr>
<td>3. DISSECTING THE BEAN BEETLE CALLOSOBRECHUS MACULATUS GERM PLASM ASSEMBLY MECHANISM IN THE CONTEXT OF TELOTROPHIC OOGENESIS</td>
<td>121</td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>122</td>
</tr>
<tr>
<td>3.2 Background</td>
<td>122</td>
</tr>
<tr>
<td>3.2.1 The maternal provision mode</td>
<td>122</td>
</tr>
<tr>
<td>3.2.2 Germ plasm assembly in Drosophila and Nasonia</td>
<td>123</td>
</tr>
<tr>
<td>3.2.3 Comparison of the ovary types</td>
<td>126</td>
</tr>
<tr>
<td>3.2.4 An evolutionary perspective on Callosobruchus germline</td>
<td>128</td>
</tr>
<tr>
<td>3.3 Methods and Materials</td>
<td>129</td>
</tr>
</tbody>
</table>
### TABLE OF CONTENTS (continued)

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.1 Callosobruchus rearing and sample collection</td>
<td>129</td>
</tr>
<tr>
<td>3.3.2 Identification and cloning of the Callosobruchus genes</td>
<td>131</td>
</tr>
<tr>
<td>3.3.3 Parental RNA interference on adult females</td>
<td>132</td>
</tr>
<tr>
<td>3.3.4 The purification of the Oskar protein for generation of polyclonal antibody</td>
<td>132</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>133</td>
</tr>
<tr>
<td>3.4.1 Both oskar mRNA and protein are localized in the assembled germ plasm</td>
<td>133</td>
</tr>
<tr>
<td>3.4.2 Oskar is required for normal oogenesis and abdominal segmentation</td>
<td>138</td>
</tr>
<tr>
<td>3.4.3 Vasa and tudor mRNAs are not localized in the oocyte but in the early embryos</td>
<td>140</td>
</tr>
<tr>
<td>3.4.4 Bruno is localized in the anterior pole of both oocyte and early embryo</td>
<td>141</td>
</tr>
<tr>
<td>3.4.5 Knockdown of bruno disrupted the localization of germ plasm and the oogenesis</td>
<td>141</td>
</tr>
<tr>
<td>3.5 Discussion</td>
<td>143</td>
</tr>
<tr>
<td>3.5.1 Markers are needed to study the functions of these genes</td>
<td>144</td>
</tr>
<tr>
<td>3.5.2 Limitations of handling Callosobruchus</td>
<td>145</td>
</tr>
<tr>
<td>3.6 Supplementary materials</td>
<td>146</td>
</tr>
<tr>
<td>3.6.1 Callosobruchus oskar cDNA sequence</td>
<td>146</td>
</tr>
<tr>
<td>3.6.2 Callosobruchus tudor cDNA sequence</td>
<td>148</td>
</tr>
<tr>
<td>3.6.3 Callosobruchus vasa cDNA sequence</td>
<td>153</td>
</tr>
<tr>
<td>3.6.4 Callosobruchus bruno cDNA sequence</td>
<td>156</td>
</tr>
<tr>
<td>3.6.5 FISH/IF protocol specific for Callosobruchus maculatus ovarioles</td>
<td>157</td>
</tr>
<tr>
<td>3.7 Literature Cited</td>
<td>160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. KNOCKING IN CALLOSBRUCHUS OSKAR GENE IN THE GENOME OF TRIBOLIUM USING CRISPR/CAS9 AND GAL4/UAS SYSTEMS</td>
<td>166</td>
</tr>
<tr>
<td>4.1 Background</td>
<td>167</td>
</tr>
<tr>
<td>4.2 Materials and Methods</td>
<td>172</td>
</tr>
<tr>
<td>4.2.1 CRISPR target site identification</td>
<td>172</td>
</tr>
<tr>
<td>4.2.2 Single guide RNA production</td>
<td>173</td>
</tr>
<tr>
<td>4.2.3 In vitro test of the sgRNAs</td>
<td>174</td>
</tr>
<tr>
<td>4.2.4 The donor plasmid productions for CRISPR_Gal4 line and UAS_CmOsk line</td>
<td>175</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS (continued)

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.4 The red flour beetle Tribolium rearing</td>
<td>175</td>
</tr>
<tr>
<td>4.2.5 Embryo collection</td>
<td>176</td>
</tr>
<tr>
<td>4.2.6 Embryonic injection</td>
<td>177</td>
</tr>
<tr>
<td>4.3 Results</td>
<td>177</td>
</tr>
<tr>
<td>4.3.1 The CRISPR_Gal4 line design</td>
<td>177</td>
</tr>
<tr>
<td>4.3.2 Single guide RNA design</td>
<td>181</td>
</tr>
<tr>
<td>4.3.3 The UAS_CmOsk line design</td>
<td>183</td>
</tr>
<tr>
<td>4.3.4 sgRNAs in vitro test showed that the Cas9-sgRNA complexes can efficiently cut the target sites</td>
<td>183</td>
</tr>
<tr>
<td>4.3.4 The in vivo injection did not result in any transgenic lines</td>
<td>186</td>
</tr>
<tr>
<td>4.4 Discussion</td>
<td>187</td>
</tr>
<tr>
<td>4.4.1 Does Tc-eagle knockout result in disruptions of embryogenesis?</td>
<td>187</td>
</tr>
<tr>
<td>4.4.2 CRISPR/Cas9 trouble-shooting</td>
<td>188</td>
</tr>
<tr>
<td>4.4.3 The possible outcomes of the cross between the two transgeneic lines</td>
<td>189</td>
</tr>
<tr>
<td>4.5 Supplementary materials</td>
<td>190</td>
</tr>
<tr>
<td>4.5.1 Additional file 1: Callosobruchus oskar coding sequence</td>
<td>190</td>
</tr>
<tr>
<td>4.6 Literature Cited</td>
<td>191</td>
</tr>
<tr>
<td>5. DISCUSSION AND OUTLOOK</td>
<td>194</td>
</tr>
<tr>
<td>5.1 New tools for deep functional analysis on the new genes we found</td>
<td>195</td>
</tr>
<tr>
<td>5.2 Identification of the protein composition in Nasonia oosome</td>
<td>195</td>
</tr>
<tr>
<td>5.3 Single cell sequencing on Nasonia pole cells, migrating germ cells, and Tribolium PGCs</td>
<td>196</td>
</tr>
<tr>
<td>5.4 Overcome the technical limitations on dissecting the germ plasm assembly mechanism in Callosobruchus</td>
<td>196</td>
</tr>
<tr>
<td>5.5 Literature Cited</td>
<td>197</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>199</td>
</tr>
<tr>
<td>APENDIX A: Permission to reuse published materials</td>
<td>200</td>
</tr>
<tr>
<td>VITA</td>
<td>212</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>1.1 Modes of germline specification: maternal provision and zygotic induction</td>
<td>3</td>
</tr>
<tr>
<td>1.2 The distribution of germ cell specification modes and oskar orthologs</td>
<td>8</td>
</tr>
<tr>
<td>1.3 Molecular Model of Oskar Function</td>
<td>12</td>
</tr>
<tr>
<td>2.1 The behavior and fate of the oosome</td>
<td>45</td>
</tr>
<tr>
<td>2.2 Genes expressed anteriorly in the pre-blastoderm stage</td>
<td>50</td>
</tr>
<tr>
<td>2.3 Transcripts localized to the oosome that are subsequently maintained in pole cells</td>
<td></td>
</tr>
<tr>
<td>2.4 mRNAs localized in the oosome (Pre-blastoderm) but are excluded from the pole cells</td>
<td>60</td>
</tr>
<tr>
<td>2.5 mRNAs strongly enriched in the posterior region of the embryos, but not enriched in the oosome</td>
<td>61</td>
</tr>
<tr>
<td>2.6 eGFP dsRNA injected embryos as the negative control</td>
<td>63</td>
</tr>
<tr>
<td>2.7 Phenotypes in Nv-bark beetle dsRNA injected embryos</td>
<td>66</td>
</tr>
<tr>
<td>2.8 Phenotypes in Nv-anillin dsRNA injected embryos</td>
<td>68</td>
</tr>
<tr>
<td>2.9 Phenotypes in Nv-rrm dsRNA injected embryos</td>
<td>69</td>
</tr>
<tr>
<td>2.10 Phenotypes in Nv-coronin dsRNA injected embryos</td>
<td>70</td>
</tr>
<tr>
<td>2.11 Phenotypes in Nv-innexin 1 dsRNA injected embryos</td>
<td>71</td>
</tr>
<tr>
<td>3.1 The schematic representation of ovarioles</td>
<td>127</td>
</tr>
<tr>
<td>3.2 The expression patterns of the cloned genes in wild type Callosobruchus embryos</td>
<td>135</td>
</tr>
<tr>
<td>3.3 The expression patterns of the cloned genes in wild type Callosobruchus telotrophic ovarioles</td>
<td>136</td>
</tr>
<tr>
<td>3.4 The oskar knockdown results on ovarioles</td>
<td>137</td>
</tr>
<tr>
<td>3.5 The oskar knockdown results on cuticles</td>
<td>139</td>
</tr>
<tr>
<td>3.6 The bruno knockdown results on ovarioles</td>
<td>142</td>
</tr>
<tr>
<td>4.1. The genome structure of Tribolium eagle</td>
<td>169</td>
</tr>
<tr>
<td>4.2. Identification of the CRISPR target sites</td>
<td>170</td>
</tr>
<tr>
<td>4.3. The expression pattern of Tc-eagle in early embryos</td>
<td>171</td>
</tr>
<tr>
<td>4.4. The3xP3-EGFP reporter shows the green fluorescence in Tribolium eye and nervous system</td>
<td>179</td>
</tr>
<tr>
<td>4.5. CRISPR/Cas9 system for targeting double-strand breaks</td>
<td>180</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>4.6. The genomic DNA fragment (TS1) for the in vitro efficiency test of the two sgRNAs</td>
<td>184</td>
</tr>
<tr>
<td>4.7. sgRNAs in vitro test showed that the Cas9/sgRNA complexes can efficiently cut the target sites</td>
<td>185</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PGC</td>
<td>Primordial germ cell</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>eRNAi</td>
<td>Embryonic RNA interference</td>
</tr>
<tr>
<td>pRNAi</td>
<td>Parental RNA interference</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>RNA Sequencing</td>
</tr>
<tr>
<td>nos</td>
<td>Nanos</td>
</tr>
<tr>
<td>gt</td>
<td>Giant</td>
</tr>
<tr>
<td>BRE</td>
<td>Bruno response element</td>
</tr>
<tr>
<td>osk</td>
<td>Oskar</td>
</tr>
<tr>
<td>aub</td>
<td>Aubergine</td>
</tr>
<tr>
<td>pgc</td>
<td>Polar granule component</td>
</tr>
<tr>
<td>glc</td>
<td>Germ cell less</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>PAM</td>
<td>Protospacer adjacent motif</td>
</tr>
<tr>
<td>MBP</td>
<td>Maltose binding protein</td>
</tr>
<tr>
<td>SgRNA</td>
<td>Synthetic guide RNA</td>
</tr>
<tr>
<td>HDR</td>
<td>Homology directed repair</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-homologous end joining</td>
</tr>
<tr>
<td>DI water</td>
<td>Deionized water</td>
</tr>
<tr>
<td>PAM</td>
<td>Protospacer adjacent motif</td>
</tr>
<tr>
<td>EGFP</td>
<td>Enhanced green fluorescent protein</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS (continued)

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nv</td>
<td>Nasonia</td>
</tr>
<tr>
<td>Tc</td>
<td>Tribolium</td>
</tr>
<tr>
<td>Cm</td>
<td>Callosobruchus</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescent in situ hybridization</td>
</tr>
<tr>
<td>IF</td>
<td>Immunofluorescence</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl decarbonate</td>
</tr>
</tbody>
</table>
SUMMARY

The germ cells are a unique group of cells that can give rise to all cell types and regenerate themselves, thereby passing genetic information from generation to generation in sexually reproducing animals. Germline establishment is a process through which germ cells are specified from zygote. Among animals, there are two major strategies that establish the germline during embryogenesis. The prevalent mode is zygotic induction, which is also the ancestral mode. The germ cells are induced by inductive signals after the zygotic genome has been activated, and usually depends on the interactions of other already specified tissue types. The other mode is considered as the derived mode of zygotic induction, which is maternal provision mode. Unlike zygotic induction mode, the germ cells are the first group of cells specified during embryogenesis by inheritance of the germ plasm, which contains germline determinants synthesized by nurse cell and deposited in the oocyte during oogenesis.

Some of the species in Holometabola use maternal provision mode, while none of insect species in Hemimetabola use maternal provision mode instead use zygotic induction, which suggests that the origin of maternal provision mode in Insecta is occurred early in the evolution of the Holometabola. For example, the honeybee Apis, the silkworm Bombyx, the red flour beetle Tribolium, etc. use zygotic induction mode. Interestingly, over the repetitive transitions between the two modes, there was no other mode involved. This suggests that the mechanisms for germline specification are labile over the course of evolution. Therefore, identification of the core regulatory network of the germline
specification that must be maintained, identification of the parts that are novel or variable, are important pieces of information to understand the evolution of the germline specification.

The process of germline specification is very diverse, even within the same mechanism. Take maternal provision as an example, the detailed process of germline development in the fly *Drosophila* is different from the wasp *Nasonia*. The *Drosophila* germ plasm is in a form of polar granules in the posterior pole of the embryo. They are small and static. However, the *Nasonia* germ plasm is named the oosome and is a large, tightly integrated, spherical structure that can move around in the posterior region of the embryos. At the beginning of pole cell formation in *Drosophila*, each of the nuclei associates with several polar granules and buds out of the embryo separately, resulting in several pole cells, whereas the *Nasonia* embryo makes a single bud and the oosome associated with several nuclei enter in this bud to pinch off the embryo and divided into several pole cells. Since *Drosophila* is a well-documented example for germline specification by maternal provision mode, identification of the genes that are involved in *Nasonia* germline development and comparison of them with *Drosophila* would help us understand what genes are conserved and what genes contribute to their unique features. In chapter 2, we used next-generation sequencing to identify the RNA components of the *Nasonia* oosome. Besides the conserved genes, we found dozens of novel genes that either do not exist or do not play roles in germline development in *Drosophila*. We characterized some of the novel genes’ functions by embryonic RNA interference and found that they all play roles in *Nasonia* germline development. Although the preliminary
functional studies did not give us details on these genes functions, they showed that the strategies that we used worked well for our purpose. Future deep functional studies are required to understand how those genes play their roles in *Nasonia* germline development.

To gain a fuller understanding of maternal provision mode, we also chose the bean beetle *Callosobruchus* to dissect the germ plasm assembly during oogenesis. Because *Callosobruchus* uses maternal provision mode and its oogenesis is telotrophic, which is different from the polytrophic oogenesis in *Nasonia* and *Drosophila*. In addition, its close related insect *Tribolium* uses maternal provision mode. Through our studies on germ plasm assembly, we will gain understanding of what genes are required for germ plasm assembly and what are lineage-specific. So far, we found that *oskar*, *vasa* and *tudor* are need for germ plasm assembly. Since we did not have proper germline markers, we were not able to observe the knockdown effects of these genes on germ plasm assembly. We also found that *Callosobruchus* bruno is localized at the anterior pole of the oocyte, which is a novelty. The specific localization might suggest an anterior role of *bruno* in *Callosobruchus*.

As I mentioned above, the germline specification mechanisms are labile over the course of evolution. Since *Callosobruchus* and *Tribolium* use different modes of germline specification, we tried to knock in the *Callosobruchus oskar* gene into the *Tribolium* genome to test how liable the two mechanisms are. I tried to knock in genes by CRISPR/Cas9, but I was not able to gain transgenic lines.
Overall my thesis was aimed at revealing the evolutionary relationships between the two modes of germline specification. We want to find out what the core regulatory network is and what genes contribute to lineage-specific features. Beside the conserved genes, we found dozens of novel genes in *Nasonia*, which could contribute to the unique process of *Nasonia* germline specification. Preliminary studies in *Callosobruchus* already showed unique features of the germ plasm assembly. More results would be expected if we are able to find suitable germline marker.
CHAPTER 1

INTRODUCTION
1.1 The process of germline development

The term germline refers to the groups of cells that are required for reproduction. In this sense, the germ cells give rise to all the cell fates and can regenerate themselves by passing genetic information from generation to generation.

In sexually reproducing animals, life begins with a zygote which is formed by the fusion of a sperm and an egg. The zygote enters embryogenesis, during which germ cells are specified by two major strategies, maternal provision mode and zygotic induction mode. In maternal provision, cells that inherit the germ plasm will become the primordial germ cells (PGCs) (Fig. 1.1). The major components of germ plasm are mRNAs and proteins, which are synthesized and transported to the oocyte to be localized during oogenesis. The process of germ plasm formation in the oocyte is called the germ plasm assembly, which will be discussed in detail in 1.2 using Drosophila and Nasonia as examples. In zygotic induction, the extraembryonic cells are induced to the PGCs with inductive signals (Fig. 1.1). The PGCs specified by maternal provision mode are the first group of cells formed before the soma, whereas those specified by zygotic induction mode can occur at various time points in embryonic (and potentially post-embryonic) development.

PGCs, as the products of the germline specification, then migrate to the developing gonad, and are sex-specifically differentiated into germline stem cells, which later enter the process of gametogenesis. In male they are termed as spermatogonia, while in female
Fig. 1.1 Modes of germline specification: maternal provision and zygotic induction. Red indicates germ plasm/germ cell markers. Gradient in middle two embryos indicates early stage of patterning with little differentiation of the soma. In maternal provision the germ cells are the first and only cells specified at this stage. The sharp borders in the bottom two embryos indicate that distinct tissues have been specified, and zygotic induction of germline fate can occur (right) [1, 9-13]. This figure is from [9].
they are termed as oogonia, or collectively they are termed as gonia. Through a series of mitotic and meiotic cell divisions, the gonia eventually become ova and spermatozoa. The life cycle again begins with the fertilization of a sperm and an egg.

1.2 Evolution of oskar and the origins of maternal provision in insects

1.2.1 Germ plasm and maternal provision mode

In 1911 Hegner showed that ablating the posterior region of freshly laid eggs with hot needles would lead to no germ cell formation in the beetle Leplinotarsa decemlineata. Together with previous experiments on removing the germ cell determinants, he proved that the granules localized in the posterior region was necessary to determine the germ cells [14].

Similar results were obtained on Drosophila. In 1931 Geigy found that UV-irradiation of the posterior pole of freshly laid eggs in led to sterile flies [22]. In 1974 Okada restored the sterility of the UV-irradiated flies by transferring the non-UV-irradiated pole plasm to the UV-irradiated flies [23]. By then it was known that the cytoplasm in the specific regions of the Leplinotarsa and the Drosophila embryos is required for germ cell formation. However, investigators were wondering if the induction of the pole cell formation by germ plasm is site-dependent or -independent. In 1974 experiments conducted by Illmensee and Mahowald showed that the “polar plasm” can induce the ectopic formation of the PGCs in the anterior region of the Drosophila embryos [24]. These classical experiments in the beetle Leptinotarsa and the fly Drosophila...
demonstrated that the cytologically conspicuous germ plasm is both necessary and sufficient to induce germ cell fate, indicating that this substance is a germ cell determinant rather than just a germ cell marker [14, 24].

1.2.2 The molecular mechanism of germ plasm assembly in *Drosophila*

Given that the germ plasm is necessary and sufficient for precocious segregation of the germ cells, elucidation of the germ plasm assembly mechanism and the germ plasm molecular components was critical for studying the maternal provision mode. Since *Drosophila* is being widely used for the studies in Genetics and Developmental Biology over a century, it is so far the best characterized model system for the molecular mechanisms of germ plasm assembly. *Drosophila* germ plasm, in the form of polar granules, is assembled at the posterior pole of the oocyte during oogenesis in a process for which the gene *oskar* is both necessary and sufficient [1, 25, 26]. During early embryogenesis, when the syncytial nuclei reach the embryo cortex, the nuclei that associate with the polar granules cellularize precociously to form the pole cells, which are the primordial germ cells of *Drosophila* and many other holometabolous insects.

*Oskar* is the central component of a regulatory network required to assemble germ plasm. Ephrussi and Lehmann showed that *oskar* induced the germ plasm assembly in *Drosophila*. They increased the *oskar* dosage in fourfold, which led to the increased amount of Vasa protein, and hence the increased amount of the germ plasm. Since the germ cell fate is determined by the intake of germ plasm, therefore the increased amount
of germ plasm resulted in the increased pole cell number from 10-15 to 40-60. They also demonstrated that oskar itself can assembly the germ plasm. The mislocalization of oskar mRNA in the anterior pole induced the ectopic formation of the germ cell at the anterior [1].

The molecular mechanism of germ plasm assembly is an oskar-directed and step-wise process. Through genetic screening, maternal genes such as cappuccino, magonashi, staufen, bruno, vasa, tudor, piwi, valois and germ-cell less, etc., were identified to be involved in germ plasm formation and localization [27]. oskar mRNA localization to the posterior pole of the embryo relies on the polarized microtubule network. Staufen and other proteins mediate the interactions between the oskar mRNA 3'UTR and the motor proteins [27, 28]. Interestingly, Kim-Ha et al. showed that the oskar mRNA 3'UTR contains distinct signal elements specific for the transport and the localization [29]. While oskar mRNA is being transported and localized to the posterior pole of the oocyte, Bruno binds to the 3'UTR sequence and repress the precocious translation before localization [30]. Once the oskar mRNA reaches to the localization region, it is released from Bruno for translation [30-35]. The Oskar protein is required to anchor the localized mRNA at the posterior pole and form a primitive complex to recruit proteins Vasa and Tudor to assemble and stabilize the germ plasm and to recruit other RNAs and proteins, such as germ cell-less and nanos mRNA which are required for germ cell differentiation and development [36-40].
The polar granules are membrane-less ribonucleoprotein granules that exhibit physical properties of phase separation. A recent paper published by the Lehmann group showed that there are two types of germ granules: the cytoplasmic germ granules that instruct PGC formation and the nuclear germ granules in the early PGC nuclei that have unknown functions [41]. Both groups of germ granules exhibit liquid and hydrogel-like phase transitioned properties [41]. The phase droplets can fuse, condense, dissolve, and exchange their protein content. The condensed liquid phase droplets transitioned to hydrogel-like phase droplets. Within the droplet, the proteins are stably associated and exchange with the environment in a low rate. They also found that the Oskar protein synergizes to promote the formation of phase transitioned condensates in the nuclei [41]. When the nuclear localization signal was removed, the nuclear germ granules also disappeared [41].

1.2.3 The mechanism of germ plasm assembly in *Nasonia*

Given the essential role of *oskar* and the evolutionary conservation of most other components of the *Drosophila* germ plasm assembly network, it was surprising that this gene seemed to be restricted to the Diptera [2, 4, 25] (Fig. 1.2). However, in 2011 Lynch et al. discovered an *oskar* ortholog in the wasp *Nasonia*, an insect in the basally branching order of Hymenoptera. They used parental RNA interference to analyze the functions of the *oskar* gene in germ plasm assembly and PGC formation. The removal of *oskar* mRNA led to defects in the polarity of the egg chamber and the oosome was not formed. *in situ* hybridization on the ovarioles and the embryos from pRNAi showed that Vasa protein is
**Fig. 1.2** The distribution of germ cell specification modes and oskar orthologs. Blue boxes indicate oskar ortholog in the genome [1-3], a box with a question mark indicates that no large scale genomic or transcriptomic data has been published to help determine whether oskar is present, those without indications appear to lack oskar based on its absence in large scale sequencing projects [4-8]. (Genus names in blue are those in which maternal provision mode is used [2, 3, 14-17]. Those in purple indicate that the presence of maternal provision has been proposed and/or independently derived [16-18]. The topology of the phylogeny is based on [8, 19]. This figure is from [9].
not localized and the nanos, otd1 mRNAs formed a homogenous cap at the posterior pole instead of an integrated oosome without pole cells formed in the later stages. These results showed that the oskar gene is required for germ plasm assembly and PGC formation, which is conserved between Nasonia and Drosophila [3].

1.2.4 The origin of oskar gene corresponds with the appearance of the maternal provision mode in insects

Besides the conserved roles of the oskar gene, Lynch et al. also showed that most of the regulatory interactions upstream and downstream of Nasonia oskar are conserved with Drosophila, implying that these two networks are homologous and share a common evolutionary origin. Since the Hymenoptera and Diptera diverged at the base of the Holometabola, and neither oskar genes, nor unambiguous maternal germ plasm had yet been found in any Hemimetabolous insects, it was proposed that oskar was a novelty of the Holometabola, and its origin more or less correlated with the origin of maternal provision in this clade [3] (Fig. 1.2). A transcriptomic analysis of the beetle Callosobruchus maculatus, which has clear germ plasm and early pole cells [42], has identified an oskar ortholog [10], further supporting the correlation between the presence of an oskar gene and the presence of maternal provision of germ plasm.

Thus, it was a major surprise when an oskar ortholog was discovered in the Hemimetabolous cricket Gryllus, suggesting that this gene has a much longer evolutionary history than previously appreciated [11]. This discovery was particularly
unexpected because *Gryllus* lacks any evidence of maternal germ plasm [12]. Indeed, strong knockdown of *Gryllus oskar* had no effect on germ cell formation, development or viability of the crickets [11]. The only detected phenotype was sporadic defects in nervous system patterning [11]. Therefore, the evolutionary origin of *oskar* appears to have predated its role in germ plasm assembly, and *oskar* was apparently co-opted for a role in germline specification in a common ancestor of the Holometabola.

The origins and molecular functions of the *oskar* gene are still somewhat mysterious, but progress has been made recently. While the origin of *oskar* has been pushed back at least into more basally branching Hemimetabolous insect lineages, it still appears to be a novelty of the insects as no *oskar* genes have been found in non-insect arthropod genomes and transcriptomes [43-47]. The discovery of *Nasonia* and *Gryllus oskar* has made it clear that there are at least two domains of Oskar (that were only weakly detectable in *Drosophila* Oskar), which are related to domains found in other molecules [3, 11]. At the N-terminus is the LOTUS domain, which is also found at the N-termini of several Tudor-domain containing genes that are highly conserved among animals [48, 49]. Toward the C-terminus there is a domain that bears some similarity to SGNH hydrolases and lipases, which was recently dubbed the OSK domain [3, 10, 50].

Two recent papers have shed light on the molecular function of Oskar. Pull down experiments using embryonic extracts showed for the first time that Oskar directly binds germline mRNAs in vivo [50]. In vitro analyses with isolated LOTUS and OSK domains showed that it is the OSK domain that binds RNA, while the LOTUS domain showed no
such activity (Fig. 1.3.) [50, 51]. Instead the LOTUS domain forms dimers, which are predicted to block any nucleic acid binding potential and mediates Oskar interaction with the RNA helicase Vasa (Fig. 1.3.) [50, 51]. This is also important, as it demonstrates for the first time a molecular function for Oskar’s LOTUS domain in *Drosophila* and provides a direct link to the downstream events of germ plasm assembly, which require Vasa protein localization [50]. In the long Oskar isoform, the N-terminal extension confers its unique function of anchoring the *oskar* mRNA and protein to the cortex of the oocyte (Fig. 1.3.) [21, 52]. In addition, the long Oskar can also entrap and maintain the mitochondria at the posterior where the PGCs are formed by binding to the actin cytoskeleton and ensures that the mitochondria can be passed to next generation through germline [53].

*Oskar* also has a more subtle role in the assembly of the germ plasm. The morphological characteristics of this substance vary among insects, from the uniform, stably anchored granules of flies, to the often large, spherical, spatially dynamic oosome of wasps [3, 54]. It has been shown that even among species in the *Drosophila* genus, germ granule morphology and size is variable [55]. Later studies showed that the differences were due to species-specific properties of Oskar proteins [56]. Changes in Oskar protein can also lead to more drastic changes in the protein’s function. For example, while Oskar from *Drosophila immigrans* is able to induce functional germ plasm formation when expressed in the ovaries of *Drosophila melanogaster*, Oskar from the equally distantly related *Drosophila virilis* cannot [57].
Fig. 1.3. Molecular Model of Oskar Function. Long and Short Oskar are synthesized at the posterior pole of the *Drosophila* oocyte. There, Short Oskar initiates the formation of germ plasm. Short Oskar interacts with the RNA helicase Vasa via its LOTUS domain (1) and binds to RNA via its OSK domain (2). Oskar’s RNA targets in the embryo include *nos*, *pgc*, and *gcl* mRNAs. Vasa might contribute to Oskar’s function in RNA metabolism (3). In addition, Oskar might regulate Vasa activity. Long Oskar contains an N-terminal extension that confers activities distinct from those of Short Oskar. Long Oskar is essential for proper anchoring of the germ plasm to the posterior cortex of the oocyte. The N-terminal extension inhibits the potential Vasa and RNA-binding activities of Long Oskar by a yet unknown mechanism. The drawing of Vasa was generated with the help of the published structure (PDB: 2DB3 [20]). This figure is from [21].
In order to gain an understanding of the above phenomena at a more detailed level, the evolutionary forces driving divergence of Oskar protein sequences among *Drosophila* species were assessed. It was found that in general, *oskar* loci are under purifying selection [58]. However, when parts of the gene are examined in isolation, patterns of positive selection are observed for specific functional regions of the Oskar protein, particularly those involved in interactions with the localization factor LASP, and the Long-OSK isoform specific sequence region [58]. It was proposed that selection on these regions might be responsible for the divergent morphology of the germ plasm granules in *D. immigrans*, and the inability of *D. virilis oskar* to rescue germ plasm in *D. melanogaster* [58].

1.3 Mechanisms of zygotic induction

Classical studies have strongly supported the idea that zygotic induction was the ancestral mode of insect germline determination, based on histological studies that consistently found a lack of maternal germ plasm, and relatively late appearance of primordial germ cells (e.g. [59, 60]). This pattern has largely been confirmed in the molecular era (e.g. [61]). More recently, the absence of maternal provision was demonstrated in the milkweed bug *Oncopeltus fasciatus* (a hemipteran member of the Paraneoptera, the sister group to the Holometabola) [62]. In this species, some germline markers are expressed maternally, but are not localized, and there is no *oskar* ortholog. Instead, the first specific localization of germline genes such as *vasa*, *tudor*, *nanos*, and *piwi* occurs late in the cellular blastoderm stage at the posterior pole, just as the first signs
of embryonic morphogenesis are taking place. Especially, vasa gene is maternally expressed across species, but only localized in the germline. So far, the molecular basis of this zygotic induction of germ cell fate is not known [62].

Recently, the zygotic induction mode of germline determination was characterized at the molecular level in the cricket *Gryllus bimaculatus*. The authors hypothesized that the role for BMP signaling in inducing germ cells might be a conserved mechanism used ancestrally by insects [63-68]. By both knocking down and overexpressing BMP signaling activity, they were able to show that indeed BMP has a crucial role in regulating the formation of primordial germ cells in the late stages of *Gryllus* embryogenesis [68]. It is not clear whether the conserved role of BMP signaling in *Gryllus* and mammals represents an ancient homology, present in their most recent common ancestor, or whether BMP was independently recruited for germline determination in these lineages.

Nakamura and Extavour demonstrated that the transcription factor Blimp-1 has a role in germline determination downstream of BMP signaling in *Gryllus* [69]. Since the Blimp-1 ortholog in mice has a similar role in specifying germline fate downstream of BMP signaling, the *Gryllus* result indicates that the gene regulatory networks for germline determination in crickets and mice share components and structure. This supports a common evolutionary origin of the zygotic induction mode in insects and vertebrates. Further analysis and comparison of the downstream gene regulatory network in currently established model species, as well as further sampling of zygotic induction mechanisms in additional basally branching insect lineages (such as cockroaches, dragonflies, mayflies and Apterygote insect lineages), will hopefully give us a clearer idea of what the
ancestral mechanism of insect germline determination was. This can then be compared with additional data from other animal lineages to test the hypothesis of shared ancestry of BMP based germline determination.

All of the mammals utilize the zygotic induction mode to specify their germline. As a mammalian representative, the mouse germ cells were identified in the posterior of the primitive streak in the extraembryonic mesoderm at 6.5 days after mating [70]. Later experiments showed that the zygotic inductive signals (Bmp4, Bmp8b and Bmp2) can induce the proximal epiblast cells and even the distal epiblast to adopt the germ cell fate [63-65, 71]. Vasa gene is also conserved in mouse, which is the mouse vasa-homolog (MVH). The MVH protein is expressed in the PGCs in the gonads during the gametogenesis in both male and female and is localized in the spermatogenic cells of the testis [72, 73].

1.4 Evolutionary liability of germline determination modes

Given that germline cells are, from an evolutionary point of view, the only cells that matter in organisms that cannot reproduce by budding or fragmentation, it seems counterintuitive that there would be much evolutionary change in their establishment. However, this is the pattern that is observed. In what we consider the current most parsimonious model of germline evolution in insects, there was at least one derivation of maternal provision at the base of the Holometabola, then at least three independent ‘reversions’ to zygotic induction associated with loss of oskar loci within the Holometabola:
in bees (*Apis*) [5, 74], tenebrionid beetles (*Tribolium*) [3, 75], possibly moths and butterflies (*Bombyx, Pararge*) [76-79], and likely more cases that have not been discovered due to the sparse sampling of insect embryonic developmental mechanisms relative to the massive diversity of Holometabola (Fig. 1.2).

Interestingly, these known ‘revertants’ to zygotic induction have adopted distinct strategies for germline establishment. The honeybee germline is established very late in embryogenesis in association with the mesoderm [60, 80, 81], which is at least superficially reminiscent to the emergence of germ cells from among the mesoderm in *Gryllus*. In contrast, *Tribolium* germ cells form at the posterior pole at the onset of gastrulation and ingress to the interior of the embryo [75], in a manner very similar to *Oncopeltus* and other hemipterans ([62], and references therein).

*Bombyx* appears to be an outlier in that the first evidence of germ cells (based on high levels of Vasa protein) appears at the late blastoderm stage in the middle of the germband [82]. More recently it was shown that a *nanos* ortholog (*nanos-O*) is distributed in granules in the ventral region of early *Bombyx* embryos, and is also found in the early germ cells in the late blastoderm stage [82]. In addition, a distinct cytoplasmic structure containing *nanos-O* RNA was found in the butterfly *Pararge* [83]. This may indicate that there is some type of maternal germ plasm in *Bombyx* and other Lepidoptera. However, given the lack of a strict spatial correlation between the *nanos-O* granules and the location of the first germ cells, and the late appearance of germ cells well after germband formation, it is premature to overturn the classical interpretations of higher Lepidopteran germ cell determination without more direct functional evidence [84].
An example of the lability of germline determination modes moving in the opposite direction may be found in the aphid *Acyrthosiphon pisum*, where a region of the early embryo accumulates the germline-associated gene *vasa* (mRNA and protein), and the cells deriving from this region appear to eventually become primordial germ cells in both sexual and asexual forms [85, 86]. Since the localization of germline markers occurs before the blastoderm forms, it is unlikely to be due to zygotic genome function, and thus could be considered a form of maternal provision that was independently derived in the aphid lineage.

This reinvention must use distinct molecular and cellular mechanisms from those known in Holometabola. First, like the other Paraneopterans so far sequenced [6, 7], *A. pisum* lacks an *oskar* ortholog [7] (Fig. 2). In addition, the mechanisms of localization must also be diverged, as the putative germ plasm only becomes localized after several syncytial divisions, appears as a ring at the cortex some distance away from the posterior pole, and is not cytologically distinguishable from the surrounding cytoplasm. Finally, aphid germ plasm does not induce precocious cellularization of the primordial germ cells [86].

Clearly caution is warranted in making strong statements about the evolutionary history of insect germline determination, as a full accounting of just how labile germline determination mechanisms have not been achieved, and sampling of species relative to the total diversity of insects is almost vanishingly small. Looking at outgroups [19, 87] is also not necessarily helpful. The myriapods [45, 88] and chelicerates [46, 89] examined thus far use zygotic induction and lack *oskar* (Fig. 1.2). However, representatives of
closely related crustacean [90] and collembolan [91] clades apparently produce maternal germ plasm, and there is even some evidence for a maternal germline determinant in the apterygote insect *Thermobia* [92] (Fig. 1.2.). Such evidence could be used to support the (unlikely, in our opinion) hypothesis that maternal provision is ancestral to insects, and that some hemimetabolans simply lost this feature.

Another mystery is the role of *oskar* in the transitions between germline determination modes. Those species that lack clear maternal provision in the Holometabola also all lack *oskar* orthologs in their genomes (Fig. 1.2.). *Oskar* is also missing from several Paraneoptera, including true bugs and aphids, yet it persists in *Gryllus* [11]. Our current understanding indicates that *oskar* arose early in insect evolution, where it initially took on a non-germline function, likely in the nervous system, and then was recruited for a germline function in a common ancestor of the Holometabola [11]. If the nervous system function is even weakly advantageous, *oskar* should be maintained in the genome even if its germline function were lost. On the other hand, if its nervous system function is superfluous, it is hard to imagine how it persisted long enough to be repurposed as a germline determinant. In addition, it cannot be formally excluded that loss of maternal provision of germ plasm could be an indirect effect of selection favoring the loss of *oskar* in another functional context. A better understanding of the non-germline functions of *oskar* in the species where it has a germline function (such as *Drosophila*, *Callosobruchus*, and *Nasonia*) could provide insights into how this gene could have been lost in members of their respective Orders.
This question of oskar’s evolutionary history leads to inquiries about the evolutionary forces driving transitions between germline modes. Are there any ecological and/or embryological conditions that favor one mode of germline determination over another? This question could be addressed by first gaining a higher resolution understanding of where in phylogeny transitions have occurred, which will, again, involve deeper sampling of insect diversity. Evans et al. found that in vertebrates, maternal provision was correlated with significant increases in both species diversity and rates of molecular evolution in maternal provisioning clades relative to zygotic inducing ones [93]. This view was updated by a recently published hypothesis paper, which argued that it is the early germline specification during embryogenesis, rather than the way germline is specified, affects the evolvability of somatic development [94]. Whether such a pattern (and its reciprocal) is found in insects is an interesting open question, which may provide a partial explanation for the pattern of change in germline determination modes across insect phylogeny.

1.5 The diversity of germline specification

1.5.1 Germline specification in Drosophila melanogaster and Nasonia vitripennis

Although there are only two major strategies to specify the germ cells, the developmental processes of the germline specification in different animals are quite diverse. Drosophila is the most well characterized model insect for maternal provision mode. The germ plasm assembled in the form of granules, also named polar granules, is localized in the posterior pole of the oocyte in the context of polytrophic oogenesis. The
polar granules are static during early embryogenesis until a single nucleus associates with some of the granules and bud out of the embryo individually at beginning of the blastoderm stage [32]. Usually few pole cells are formed at the posterior pole of the embryo, and then as the embryo develops to late blastoderm stage the initial pole cells are divided into more pole cells which are ready to migrate to the gonad [1].

As an insect from a basal branching order in Hymenoptera, *Nasonia* also uses the maternal provision mode to establish its germline. Although the regulatory network for germ plasm assembly in *Nasonia* is very similar to that in *Drosophila*, the process of pole cell formation is different. Like *Drosophila*, the *Nasonia* germ plasm is also assembled and localized in the posterior pole of the oocyte in the context of polytrophic oogenesis. However, the morphology of the *Nasonia* germ plasm is a large, integrated, spherical structure named oosome that can move freely in the posterior region of the embryo during early embryogenesis [3]. The furthest distance the oosome moves away from the posterior pole is almost near the midline of the embryo. Before budding, the oosome moves to the posterior pole and becomes hollow. While the embryo is budding, the hollow oosome flattens as a thin layer and attaches to the cortex of the bud. The bud will quickly divide into three to five initial pole cells, which will continue to divide into 24 to 28 pole cells by the late blastoderm stage (Lynch and Quan, unpublished).

Finally, the pole cells have different behaviors during gastrulation. The *Drosophila* pole cells remain on the posterior surface and migrate to the gonad [95], whereas the *Nasonia* pole cells ingress into the blastoderm cells at the onset of gastrulation [3].
1.5.2 Germline specification in *Caenorhabditis elegans* and *Danio rerio*

The germline of nematode *C. elegans* is specified by maternal provision mode. The electron-dense P granules in *C. elegans* like *Drosophila* polar granules are assembled during oogenesis stage. However, unlike the polar granules, the P granules are not localized in the oocyte before pronuclear fusion [96]. The P granules then move to the posterior of the embryo and are all segregated to P4 cell after 4 asymmetrical cell divisions. As the first PGC, P4 cell then divides into two daughter cells Z2 and Z3, which begin to symmetrically divide into more PGCs at the end of first larval stage [97]. The *C. elegans* does not have the *oskar* homolog, but it has a gene named *deps-1*, whose protein accumulates mRNAs such as the Vasa-like RNA helicase *glh-1* mRNA and promote the germ granules assembly [98].

Zebrafish is a vertebrate representative that uses maternal provision mode. The maternal mRNA, *vasa*, first appears in the 1-cell stage at the cleavage furrow [99, 100]. At 32-cell stage, the mRNA becomes four subcellular clumps, which eventually are segregated into four PGCs during 1k-cell stage. These four PGCs start to divide to more PGCs at 4K-cell stage [101]. The *oskar* homolog does not exist in zebrafish, instead the gene named *bucky ball* play a similar role in organizing the germ plasm assembly [102].

1.6 The evolutionary necessity of the chosen insects
*Drosophila* is the most well characterized model organism for germline development, specifically the maternally provided germline specification. However, for evolutionary studies, *Drosophila* only represents a small order of insects in Diptera, which is highly derived and lineage specific. The wasp *Nasonia*, on the other hand, is a representative of the basally branching in insects in Hymenoptera. In addition, both *Nasonia* and *Drosophila* use maternal determinants to specify their germline. Therefore, comparisons between the lineage specific insect and the basally branching insects are important for understanding the evolution of the maternal provision germline specification mechanism. Indeed, although they share a lot of essential features during germline establishment, they both have their own unique features during this process, which raise questions such as what the components of the core regulatory network are for this type of germline specification and what genes contribute to their uniqueness?

Two Coleopteran insects, *Callosobruchus maculatus* and *Tribolium castaneum*, are also of interest. Although the bean beetle *Callosobruchus* uses maternal provided determinants to establish its germline as well, it is quite different from *Nasonia* and *Drosophila*. As we discussed above, the germ plasm is assembled during oogenesis. Therefore, the mode of the oogenesis might affect the germ plasm assembly. Unlike the polytrophic oogenesis in *Nasonia* and *Drosophila*, *Callosobruchus* has telotrophic oogenesis as *Tribolium*. In telotrophic oogenesis, a common population of nurse cells are found at the anterior tip of the ovarioles, and the youngest oocyte is the closest one to the nurse cells. Each of the oocytes is connected to the nurse cells via a long projection called trophic cord. In contrast, there is a group of nurse cells located at the anterior side
of the oocytes in polytrophic oogenesis. This raises questions like, how the germ plasm is assembled in the context of the telotrophic oogenesis and is this mechanism any different from the that in the polytrophic oogenesis? We will try to answer these questions in Chapter 3.

A population of posterior cells are induced by signals to be the germ cells in the red flour beetle *Tribolium*, and there are no signs of assembled germ plasm found so far. Interestingly, its closely related beetle *Callosobruchus* uses maternal provision mode, even though they both have the telotrophic ovarioles. This put the liability of the germline specification mechanism into perspective, since the ancestral mode is zygotic induction, while the maternal provision mode is a derived mode for a limited number of animals. Examination of the phylogenetic tree of the insects, *Tribolium* is one of the insects that lost the ability to use maternal provision mode after it was gained by the common ancestor of Holometabola. This phenomenon also appears in the honey bee *Apis* and the silkworm *Bombyx*. Therefore, as an emerging insect that is closely related to *Callosobruchus*, *Tribolium* serves well for the evolutionary questions on germline establishment, such as how conserved the regulatory network is between the two insects, and by extension how conserved the regulatory network is between the two modes among all the sexually reproducing animals.

By comparing the three insects using maternal provision mode (*Drosophila, Nasonia, Callosobruchus*), we are expecting to identify the genes that are conserved for this mode of germline specification, the genes that are lineage specific, and the
differences of the germ plasm assembly mechanisms between polytrophic and telotrophic oogenesis. By comparing the studies on these four insects (*Drosophila, Nasonia, Callosobruchus, Tribolium*), we are expecting to identify the conserved regulatory network between the two germline specification modes.

1.7 Thesis goals and organization

The overall goals of this thesis were to identify the core regulatory network for both the germline specification modes, and the novelties that contribute to the mode specific and the lineage specific germline development. In addition, to reach to these goals it was also necessary to develop tools for these emerging non-model insects.

In chapter 2, I examined the RNA components of the *Nasonia* oosome using the next-generation sequencing. I compared the localized RNAs to that of the polar granules in *Drosophila* and found a handful of conserved genes and dozens of novel genes. Functional analysis was also performed to gain preliminary knowledge of these genes in germline development.

In chapter 3, I dissected the germ plasm assembly in the telotrophic oogenesis in the beetle *Callosobruchus*. I cloned the *Callosobruchus* orthologs of the genes localized in the *Nasonia* oosome and the *Drosophila* polar granules. I also examined their expression patterns in the oocyte and their knockdown effects on germ plasm assembly and oocyte maturation as well as later embryonic and germline development. To perform
these studies, I optimized fluorescent in situ hybridization (FISH) protocol specific for *Callosobruchus* ovarioles and purified the full-length Oskar protein to make the polyclonal antibody. I also identified several germline markers. Surprisingly, the *bruno* mRNA is expressed anteriorly in the *Callosobruchus* oocyte, which makes a good anterior marker.

In chapter 4, I tried to prove the liability of the germline specification mechanisms between the two modes and examine the role of *oskar* gene in the transitions between the two modes in Hymenoptera. To achieve these goals, I tried to knock in the *Callosobruchus oskar* gene into its close relative, the red flour beetle *Tribolium*, for maternal expression using the CRISPR/Cas9 and the Gal4/UAS systems.

In the last chapter, I summarize the progress of my Ph.D. work and highlight the important contributions to this broad topic of research. I also proposed future projects that potentially will further this work and increase our understanding of the evolution of insect germline specification.

1.8 Literature Cited


   Ultrastructural study on the origin of primordial germ cells. *Journal of embryology

17. Lin GW, Cook CE, Miura T, Chang CC: Posterior localization of ApVas1 positions
   the preformed germ plasm in the sexual oviparous pea aphid Acyrthosiphon

18. Rost MM, Flakus A, Klag J: Primordial Germ Cell Differentiation in Natural and
   Manipulated Twin Embryos of Thermobia domestica (Insecta: Zygentoma).

   Ware J, Flouri T, Beutel RG *et al*: Phylogenomics resolves the timing and pattern

    for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* 2006,
    125(2):287-300.

    Structure of the Drosophila Germline Inducer Oskar Identifies Two Domains with
    Distinct Vasa Helicase- and RNA-Binding Activities. *Cell reports* 2015, 12(4):587-
    598.

22. Geigy R: Action de l'ultra-violet sur le pôle germinal dans l'oeuf de Drosophila


CHAPTER 2

TRANSCRIPTOMIC AND FUNCTIONAL ANALYSIS OF THE OOSOME, A UNIQUE FORM OF GERM PLASM IN THE WASP NASONIA VITRIPENNIS
2.1 Abstract

2.1.1 Background

The oosome is the germline determinant in the wasp *Nasonia vitripennis* and is homologous to the polar granules of *Drosophila*. Despite a common evolutionary origin and developmental role, the oosome is morphologically quite distinct from polar granules. It is a solid sphere that migrates within the cytoplasm before budding out and forming pole cells.

2.1.2 Results

To gain an understanding of both the molecular basis of the novel form of the oosome, and the conserved essential features of germ plasm, we quantified and compared transcript levels between embryo fragments that contain the oosome, and those that did not. The identity of the localized transcripts indicate that *Nasonia* uses different molecules to carry out conserved germ plasm functions. In addition, functional testing of a sample of localized transcripts revealed potentially novel mechanisms of ribonucleoprotein assembly and pole cell cellularization in the wasp.

2.1.3 Conclusions

Our results demonstrate that numerous novel and unexpected molecules have been recruited in order produce the unique characteristics of the oosome and pole cell
formation in *Nasonia*. This work will serve as the basis for further investigation into the patterns of germline determinant evolution among insects, the molecular basis of extreme morphology of ribonucleoproteins, and the incorporation of novel components into developmental networks.

2.2 Background

Germline establishment is a crucial event for sexually reproducing organisms. Germline cells are special in that they are able to generate all of the cell fates of the soma and to regenerate themselves. There are two major strategies to specify the germline among animals: zygotic induction and maternal provision. In zygotic induction, inductive signals from surrounding tissues drive the establishment of germline fate, usually relatively late in embryogenesis, after the transition from maternal to zygotic control [1]. In contrast, in the maternal provision mode, the germ cells are specified by determinants called germ plasm that are synthesized and localized during oogenesis and are the first group cells formed during embryogenesis. Germline specification in this mode occurs very early in development, usually prior to the activation of the zygotic genome [1, 2]. Classical experiments have shown that germ plasm is both necessary and sufficient to establish the germline fate [3-5].

It is likely that the maternal provision mode of germ plasm evolved multiple times among animals, and this is reflected in the molecular basis of germ plasm determinants [6-27]. In vertebrates, germ plasm (where it exists) is dependent on the maternal
localization of bucky ball [28]. In insects, the gene products of oskar (osk) are both necessary and sufficient to induce germ plasm, and thus, primordial germ cells (PGCs) [29-32]. Downstream of these nucleators is a suite of highly conserved germline-associated molecules (i.e., Vasa (Vas), Nanos (Nos), Tudor (Tud), etc.) that are recruited to the germ plasm where they presumably carry out functions ancestral for the animal germline [30, 33-35].

There are several conserved properties of PGCs that must be conferred to naive cells by germ plasm. One of these is a period of transcriptional quiescence germ cells undergo after being specified. This feature makes the composition of a germ plasm even more important since it means that mRNAs and proteins critical for at least the early germ cell functions need to be provided in the germ plasm itself. Germ cells also usually become highly migratory later in development as they seek to colonize the developing gonad. They are often highly enriched for mitochondria and have specific metabolic needs [36]. Since they carry the genome that will be passed to future generations, germ cells have enhanced mechanisms to prevent DNA damage and to reduce the activity of transposable elements [37, 38]. Finally, they must have the capability to induce pluripotency to their genome, as germline cells will be the source for all cell fates in the eventual progeny.

Like all developmental processes, there is likely to be variation in the details of these conserved features of germline determination, whether due to selective or neutral forces. Pressures that could impact the composition of the germ plasm could be
differential activity of transposable elements in the genome, or a novel path for migration of the PGCs to the gonad, for example. Novelties in embryogenesis may also drive germ plasm composition. For example, in the parthenogenetic, paedomorphic embryos of the midge *Miastor*, somatic nuclei undergo significant chromosomal diminution in the early cleavages while the cell that takes up germ plasm maintains the full complement of chromosomes, indicating that the germ plasm contains a component that prevents the loss of chromosomes during the early cleavages [39].

Holometabolous insects are an ideal system with which to study how germ plasm evolves, given that it is ancestral in this clade [35], the unparalleled levels of diversity, and the strong baseline understanding of germ plasm function derived from *Drosophila melanogaster*. Here we focus on the parasitic wasp *Nasonia vitripennis* as a model to compare to the fruit fly. Like *Drosophila*, *Nasonia* depends on *oskar*, *vasa*, and *tudor* to assemble germ plasm [35, 40]. However, in contrast to the collection of small granules stably associated with the posterior pole that make up the *Drosophila* germ plasm, the *Nasonia* germ plasm forms a very large, dense organelle called the oosome (Fig. 2.1). This highly divergent morphology strongly indicates that the composition of the *Nasonia* oosome may diverge significantly from the polar granules of *Drosophila*.

The behaviors of the oosome and the PGCs in *Nasonia* further imply a divergent composition of the oosome. In freshly laid eggs, the oosome is tightly bound to the ventral-posterior cortex of the embryo (Fig. 2.1 A). When the zygotic nucleus forms and moves into the interior of the embryo, the oosome detaches from the cortex and coalesces into
a dense, extremely large structure in the same central column as the nuclei. It migrates anteriorly, sometimes being found near to 50% egg length, before migrating back to the posterior pole (Fig. 2.1 B-D). As the cleavage nuclei migrate toward the cortex, the oosome flattens into a crescent on the posterior pole of the embryo while a large bud protrudes from the pole (Fig. 2.1 E). Typically, two or three nuclei become associated with the bud and the oosome material. The bud pinches off, and the nuclei rapidly individuate into pole cells (Fig. 2.1 F, G).

This is distinct from pole cell formation in *Drosophila*, where each *Nasonia* pole cell forms individually upon association with a certain critical amount of pole plasm [41]. After dividing a few times, the pole cells remain stable at the posterior pole of the egg until gastrulation, when they migrate through the posterior epithelium and coalesce into to two masses, presumably where the primordial gonads are developing (Fig. 2.1 H). This migration is distinct from *Drosophila* pole cells which remain external on the posterior pole as the germ band undergoes massive extension well after gastrulation and migrate internally through the gut wall [41].

Thus, it is clear that *Nasonia* and *Drosophila* share some fundamental aspects of germline establishment, but they also have their own diverged features that fit in their own embryogenesis program. This raises the question of which genes are the core components for maternal provision mode and which genes contribute to their own distinct features in germline development.
Fig. 2.1. The behavior and fate of the oosome. Green marks the germ plasm (Nv-bark in situ hybridization), white (DAPI) marks the nuclei. A-D: the green shows the oosome shapes and localizations in different nuclear divisions during pre-blastoderm stage. E: the oosome flattens into the bud (green) as the nuclei reach the embryonic cortex. F: several pole cells immediately form from the bud during early syncytial blastoderm stage. G: the pole cells divide several times to make more and smaller cells by the time of cellularization. H: the pole cells are internalized during gastrulation and begin to migrate to the where the gonads will form. Anterior is at left, and dorsal is up in all figures.
To address these questions, we compared the mRNA content of anterior and posterior halves of the pre-blastoderm stage *Nasonia* embryos in an effort to identify the components specifically localized to the oosome. We found only a few of mRNAs conserved in both *Drosophila* polar granules and the *Nasonia* oosome, such as *oskar*, *nanos*, and *ovo*. The rest are all novel components of germ plasm that either lack *Drosophila* homologs or have homologs in *Drosophila* that do not play any roles in germline development. We performed functional studies for a set of localized transcripts, all of which showed roles in the unique features of *Nasonia* germline development, demonstrating the value of our approach to identify the molecular sources of novelty among various insect lineages.

2.3 Results

2.3.1 RNA-seq analyses of the anterior and posterior poles of the wasp *Nasonia* early embryos

To identify the maternal transcripts in the oosome, we isolated the total RNA separately from anterior and posterior poles of the pre-blastoderm stage *Nasonia vitripennis* embryos. Six samples that comprised three sets of each pole were sequenced and subjected to differential expression analyses, which are described in more detail in the Methods.

Our first attempt yielded encouraging, but mixed results. Several novel posteriorly enriched factors were identified, and we did recover some known posterior factors (*Nv-
oskar and Nv-nanos). However, other known localized factors such as Nv-cad and Nv-dpp were not found to be statistically significant. In addition, one of the posterior samples was of poor quality, and could not be used in the differential expression analyses, reducing the power of our approach. Since we wanted results as comprehensive as possible, we repeated the experiment using a few adjustments (see Methods) and obtained much more robust results. While this second experiment was being prepared, we moved forward with expression and functional analysis of novel candidates identified in the first. Since a handful of transcripts with confirmed oosome localization were found only in the first experiment, the remainder of the results will include these transcripts along with those identified in the second experiment.

All known maternally localized molecules (Nv-osk, nos, dpp, cad, otd1, and gt) were found with high significance, giving confidence that our experimental design and analysis approaches were appropriate for our goal of discovering the mRNAs localized to the oosome. Overall, we found 92 transcripts with apparent significant enrichment at the posterior pole. These ranged in levels of enrichment from 1.4 to 55 times higher in the posterior fragments compared to anterior fragments. Our analyses also uncovered anteriorly enriched mRNAs, of which there were 89, with a range of fold enrichment from 1.4 to 10 times higher at the anterior.

2.3.2 Novel transcripts localized in the anterior half of the Nasonia early embryos
While the anterior factors are not the focus of this manuscript, some interesting observations were made in examining a small sample of candidate mRNAs. Most transcripts are localized in small domains at the anterior cortex and seem to extend toward the posterior in variable tendrils, rather than being uniform or graded caps of anterior localization (Fig. 2.2 A1-K1). A notable exception is the transcript of *Nasonia* homolog of *mex-3*, an RNA binding protein known for controlling translation of orthologs of the posterior patterning factor *caudal* in the nematode *C. elegans* and the beetle *Tribolium* [24, 42]. *Nv-mex3* mRNA is localized in a broad domain extending far toward the posterior of the embryo (Fig. 2.2 L1). *Nv-mex3* expression is highly dynamic and variable in both the blastoderm and post-gastrular stages (Fig. 2.2 L2, L3, additional images not shown).

Transient localization is the most common feature of the anteriorly localized transcripts. Most are ubiquitous or absent by the time the early syncytial blastoderm forms (Fig. 2.2 A2-H2, J1-K3, M1-M3), except for *LOC100313502* which persists into the blastoderm stage where it forms an anterior cap (Fig. 2.2 I1-I3). Among the anteriorly enriched transcripts, many have predicted functions that may be relevant to egg activation (e.g. four encoding ion channels), and anterior-posterior polarity and patterning (e.g. transcription and translation factors, see 2.7.1). There are also a large number of transcripts with no clear homologs outside of *Nasonia* see (2.7.1). One of these (*LOC100119982*, Fig. 2.2 G1-G3) is a member of a novel family of ankyrin domain containing molecules that are specific to Chalcid wasps and appear to have obtained a broad diversity of expression and potential function during *Nasonia* development [43].
Finally, a handful of molecules localized at both poles were detected. These include the known factors *Nv-otd1* [44] and *Nv-TollC* (JAL in preparation), along with *Nv-endoglucanase* (Fig. 2.3 E), *Nv-insulin-like growth factor* (*Nv-igf*, Fig. 2.3 C), *Nv-ucth* (a ubiquitin carboxy terminus hydrolase, Fig. 2.2 M)). The fact that these were found despite lower apparent fold differences between the two embryonic halves, further gave confidence that our analysis was robust enough to detect even subtle germ plasm localization of the vast majority of mRNAs.

2.3.4 General description of the novel transcripts localized in the posterior pole of the *Nasonia* early embryos

From the two analyses, we identified 92 candidate transcripts that were statistically enriched at the posterior half of the *Nasonia* embryos. We then isolated PCR products for 54 of these genes and made probes to determine their expression patterns during early embryogenesis. We were able to confirm 47 transcripts that are expressed posteriorly during pre-blastoderm stage. Of the remaining transcripts, one of them was not successfully cloned, and the rest were successfully cloned but did not have any specific, localized expression. Interpretation of these potential false positives (as well as some false negatives) will be discussed later. Among the 47 transcripts, 28 transcripts with expression in the pole cell formation and beyond were identified. We consider these to be the strongest candidates for having important roles in the specification and function of the PGCs but cannot exclude *a priori* that they have other (or no) important functions. Six
Fig. 2.2. Genes expressed anteriorly in the pre-blastoderm stage. All embryos are aligned and grouped into three columns (Pre-blastoderm, Blastoderm and Gastrulation) according to their embryogenesis stages, with posterior side to the right and dorsal side on the top.
transcripts are maintained throughout the stages of germline development followed in this manuscript: oosome, pole cells, and migrating germ cells. They include *Nasonia* homologs of the fly genes *bark beetle* (*Nv-bark*) (Fig. 2.1), *spt5* (*Nv-spt5*), *tejas* (*Nv-tdrd7*), *insulin-like growth factor* (*Nv-igf*), and two transcripts without fly homologs (*Nasvi2EG001470* and *Nv-endoglucanase*) (Fig. 2.3 A1-E3). Among these, *Nv-bark* is the best germline marker, bearing strong and consistent germline association throughout early development, including expression in the late embryonic gonads (Fig. 2.1). In this respect, it is better than our previously favored marker, *Nv-nos*, which is downregulated significantly toward the end of PGC migration [40]. Bark is a large transmembrane protein and is not expressed in the *Drosophila* germline. Its only known role is in stabilizing tricellular junction in epithelial cells during embryogenesis [45, 46]. It is not clear how this function is relevant to germline function, and indicates a novel recruitment of this factor in the wasp.

*Nv-spt5* is significantly enriched in the oosome with low levels of ubiquitous expression in the rest of the embryos in pre-blastoderm stage (Fig. 2.3 A1). In the blastoderm and migrating germ cell stages expression is reduced but still enriched in the pole cells while the ubiquitous expression in the embryo persists (Fig. 2.3 A2-A3). Spt5 homologs are involved in regulating RNA polymerase progression during transcription [47], which might indicate that Nv-Spt5 is involved in repressing or otherwise regulating the onset of transcription in the germ cells.
*Nv-ttdrd7* is present at appreciable levels throughout the bulk cytoplasm and is also strongly localized in the oosome (Fig. 2.3 B1). This pattern is well reflected in the quantification of mRNA levels in the two halves of the embryo, which show significant numbers of reads coming from the anterior half of the embryo. At the blastoderm stage, *Nv-ttdrd7* is moderately enriched in the pole cells and is zygotically expressed in a ventral-posterior patch (Fig. 2.3 B2), which was detected in our earlier analysis of dorsal-ventral patterning [48]. After gastrulation, *Nv-ttdrd7* is strongly upregulated in a group of cells that are near to where the germ cells migrate, but it is not clear if they are germ cells (Fig. 2.3 B3). *Nv-igf* is initially expressed in a bipolar pattern (similar to *Nv-otd1* [44]), before becoming specific to the pole cells during the blastoderm stage and the migrating germ cells after gastrulation (Fig. 2.3 C1-C3). *Nasvi2EG001470* encodes a short peptide of 80 amino acids and was not included in the most recent annotation of the *Nasonia* genome at NCBI, but was present in OGS 2.0 [49]. A very similar sequence is annotated in the close relative *Trichomalopsis* (JAL personal observation), indicating that it is a bona fide transcript that is either novel, or very rapidly evolving. *Nasvi2EG001470* is strongly expressed in the oosome and pole cells, while levels markedly decrease in migrating germ cells (Fig. 2.3 D1-D3).

Besides the expression in the oosome and the pole cells, *Nv-endoglucanase* is initially expressed at both poles during pre-blastoderm stage and early blastoderm stage (Fig. 2.3 E1-E2). Later in the blastoderm stage, the expression is down-regulated at the anterior pole and become specific to the pole cells (Fig. 2.3 E3). Proteins of this type are found extensively in Hemimetabola, Hymenoptera, and Coleoptera, and appear to have
Fig. 2.3. Transcripts localized to the oosome that are subsequently maintained in pole cells. All embryos are aligned and grouped into three columns (Pre-blastoderm, Blastoderm and Gastrulation) according to their embryogenesis stages, with posterior side to the right and dorsal side on the top.
been lost at the origin of Lepidoptera and Diptera clades. It would be intriguing to determine what roles this protein plays during Nasonia embryogenesis.

Five transcripts localized to the oosome enter and are maintained in the pole cells but are then downregulated in the migrating germ cells (Fig. 2.3. F1-J3). This set includes homologs of Drosophila anillin (scraps), qin, and innexin1 (ogre), (Nv-anillin, Nv-qin, and Nv-innexin1, respectively). Anillin is an actin binding protein that localizes to the contractile ring during cytokinesis [50]. In Drosophila Anillin protein is localized in the cleavage furrows when forming the PGCs [51], but its mRNA is ubiquitous. The early oosome localization of Nv-anillin mRNA suggests that the protein might also play roles in oosome outside of a potential conserved role in pole cell formation.

Nv-qin encodes a protein containing tudor domains along with an E3 ubiquitin ligase domain. qin is important in processing germline piRNAs, repressing retroelements assembling the nuage, and proper completion of oogenesis in the fly [52-54]. While qin has an important late role in germline cells, it is only weakly and diffusely expressed during embryogenesis in Drosophila [55]. Its mammalian homolog RNF17 is required for production of specific particles in the germline nuage and for sperm development, but not for early germline specification [56].

Nv-innexin1 encodes a putative gap junction protein whose fly homolog is most well-known for its role in proper development and function of the nervous system [57]. Other unexpected roles for innexin proteins have been described and proposed in insects.
[58], but at the moment the potential functional significance of the germline localization in *Nasonia* is unclear.

Two transcripts localized to the oosome and preserved in the pole cells do not have clear *Drosophila* homologs (Fig. 2.3 I1-J3). One of these (*Nv-C16orf87*) encodes a homolog of the human protein *C16orf87* and is expressed in the posterior region as well as specifically in the oosome in pre-blastoderm stage (Fig. 2.3 I1-I3). The protein of *Nv-C16orf87* belongs to the uncharacterized protein family UPF0547, which contains the zinc-ribbon motif, and functions of this protein and its homologs are not known. It appears that this gene has been lost specifically in the Brachyceran fly lineage as it is found in beetles, moths, and some mosquitos, but not *Drosophila* (JAL personal observation).

Finally, another ankyrin domain encoding transcript is strongly localized to the oosome and is taken up into pole cells (Fig. 2.3 J1-J2). It later has a complex and dynamic pattern in the blastoderm stages (Fig. 2.3 J2-J3, additional images not shown). This transcript is a member of the newly described CLANK (Chalcid Lineage-specific ANKyrin-domain gene) family, of which there are nearly 200 in the *Nasonia* genome [43]. To differentiate from the others, we name it *Nasonia vitripennis Oosome CLANK* (*Nv-OoCLANK*).

2.3.5 **Transcripts enriched in the oosome but excluded from pole cells**
A set of 12 genes is expressed in the oosome but not transported to the pole cells. We predicted that these would have germline roles primarily in the oosome itself, or in the early stages of pole cell formation. We also expect many transcripts in this set will have roles outside of germline production, such as in embryonic patterning (as already known for *Nv-dpp* and *Nv-cad*) [59, 60]. Examples of what we consider potential embryonic patterning factors include: a CLIP protease encoding message (*Nv-mp1*) related to fly Melanization Protease and Easter (Fig. 2.4 A1-A3), *Nv-kayak* (Fig. 2.4 B1-B3) encoding at transcription factor downstream of JNK signaling [61], and *Nv-elbow* (Fig. 2.4 C1-C3) encoding a single zinc-finger transcription factor [62]. *Nv-kayak* is later expressed in a dorsal domain, indicating a conserved role in extraembryonic patterning. Interestingly, *Drosophila elbow* interacts with *orthodenticle* in specifying the ocelli and in photoreceptor cell fate determination [63], and an intriguing possibility is that these proteins work together in posterior patterning, with *Nv-elbow* possibly playing an important role in differentiating the posterior targets of *otd* from the anterior ones.

Several oosome resident transcripts have suggestive functional annotations. For example, the *Nasonia coronin* gene (*Nv-coronin*, (Fig. 2.4 D1-D3)) encodes a protein whose homologs are known to bind and modulate actin, provide links between the actin and microtubule cytoskeletons, and regulate endo- and exocytosis in several developmental contexts [64, 65]. A germline role for the *Drosophila* Coronin ortholog has not been observed. *Nasonia milton* (*Nv-milton*) is another exciting transcript (Fig. 2.4 E1-E3). *Drosophila* Milton is an adaptor protein that allows mitochondria to be loaded onto, and transported by, microtubule motors [66].
An oosone resident mRNA encodes a protein whose fly ortholog is uncharacterized, but whose function may be relevant to oosome function. This is the *Nasonia* homolog of CG4552 (*Nv-CG4552*), which encodes a protein with a TBC25 domain (Fig. 2.4 F1-F3). Proteins with this domain interact with Rabs to regulate membrane trafficking and dynamics. Such activities have been shown to be crucial for Oskar function in the fly [67], and *Nv-CG4552* may play a supporting role in regulating membrane dynamics in the wasp. Another suggestive localized factor does not have clear orthologs outside of the Hymenoptera, but it does have two predicted RNA Recognition Motifs, therefore we name it *Nv-rrm* (Fig. 2.4 G1-G3). RRM domains bind RNA and are components of proteins that regulate RNA localization and translation. This novel lineage specific protein could therefore be involved in the localization of specific RNAs in the oosome, or the regulation of translation of specific RNAs within it.

Many of the oosome localized transcripts do not have annotations that lead to simple hypotheses about their roles in specifying the germline. One of these is *Nv-waterproof*, which encodes a fatty acyl-CoA reductase. *Drosophila waterproof* produces the hydrophobic molecules that coat the tracheal tubes during *Drosophila* embryogenesis and is essential for gas filling of the trachea [68]. The protein’s novel role in *Nasonia* germline is not clear and worth to investigate, as is its early maternal expression in the oosome and later zygotic expression as the dorsal strip and in the extraembryonic tissue. (Fig. 2.4 H1-H3). *Nv-CG42269* encodes a predicted organic ion
transporter protein whose *Drosophila* homolog (CG42269) has no described function (Fig. 2.4 I1-I3).

Three oosome localized transcripts have no clear homologs in *Drosophila* or in other model organisms. *LOC100123551* has a sterile alpha motif (SAM) domain, which might indicate protein-protein or -RNA interactions (Fig. 2.4 J1-J3). *LOC103315681* contains weak similarity to the N-terminal domain of Folded-gastrulation proteins (but is not a folded gastrulation ortholog) (Fig. 2.4 K1-K3), while *Nasvi2EG022821* has no discernible conserved domains (Fig. 2.4 L1-L3). The functions of these factors will be the object of future investigation.

### 2.3.6 Transcripts enriched in the posterior pole but not specifically the oosome

There are six transcripts that are significantly enriched in the posterior region of the embryo, but do not show significant association with the oosome (Fig. 2.5). The early embryonic expression for these transcripts appears as a cap or broad posterior to anterior gradient of mRNA. The significance of such transcripts to oosome assembly or to germ cell formation is not clear. Two of this class are known developmental transcription factors: orthologs of *Zerknuellt* [69], and *Mothers against dpp* [70] (*Nv-zen* (Figure not shown, see 2.7.1) and *Nv-mad2*), respectively (there are two mad paralogs in *Nasonia*). Two transcripts are predicted to encode catalytic enzymes: a choline kinase homologous to the *Drosophila* CG2201 (*Nv-CG2201*), and a homolog of the ADP ribosylation factor-like 4 protein (*Nv-ARL4*). Finally, *LOC100680197* and *LOC100677932* have no identifiable
homologs outside of Hymenoptera. *LOC100680197* encodes a protein with MYND-type zinc-fingers and a p27-like domain, while *LOC100677932* has no clear conserved or functional domains.

2.3.7 Functional analysis by parental RNA interference showed low phenotypic penetrance

While localization of an mRNA to the oosome and pole cells may strongly suggest a function related to PGC specification, demonstration of any such function is required. We chose a sampling of five promising molecules for in depth functional analysis (*Nv-bark, Nv-anillin, Nv-rrm, Nv-coronin*, and *Nv-innexin1*).

We initially tried to apply our parental RNAi (pRNAi) approach [44], but quickly found that this was not the ideal approach. Most dsRNAs caused partial sterility, with most of the obtainable eggs being apparently normal escapers. Eventually, we managed to collect embryos with phenotypes at a very low penetrance (2%-6%) from three of the five genes that we studied with the dsRNA concentration of 1ug/ul to 2.5ug/ul. These genes are *Nv-rrm, Nv-coronin*, and *Nv-innexin1*. They all either have no pole cells formed or fewer pole cells formed, with disorganized germ plasm residue at the posterior pole of the embryos in blastoderm stage. Infrequently, *Nv-coronin* knockdown embryos were characterized by pole cells that did not migrate to the gonad, but instead remained at the pole after gastrulation.
Fig. 2.4. mRNAs localized in the oosome (Pre-blastoderm) but are excluded from the pole cells. All embryos are aligned and grouped into three columns (Pre-blastoderm, Blastoderm and Gastrulation) according to their embryogenesis stages, with posterior side to the right and dorsal side on the top.
Fig. 2.5. mRNAs strongly enriched in the posterior region of the embryos, but not enriched in the oosome. All embryos are aligned and grouped into three columns (Pre-blastoderm, Blastoderm and Gastrulation) according to their embryogenesis stages, with posterior side to the right and dorsal side on the top.
This issue was more serious for *Nv-bark* and *Nv-anillin*. At concentrations from 1.5ug/uL-2.5ug/uL, there were no eggs laid. At lower concentrations (250ng/uL, 500ng/uL and 750ng/uL), the eggs we collected all showed normal development. Since phenotypic embryos were either completely absent or extremely rare for our genes of interest, it became necessary to develop a new technique to assess the functions of the novel oosome genes we discovered.

### 2.3.8 Development of an embryonic injection protocol for *Nasonia* RNAi

To circumvent the low penetrance problem from the pRNAi, we developed a protocol for embryonic injection of dsRNA followed by fixation and *in situ* hybridization (see details in Methods and Materials). As a negative control, we injected dsRNA against eGFP to test whether the physical injection and the dsRNA itself would affect the structure of the oosome and formation of the pole cells non-specifically (Fig. 2.6). We were happy to find that even in embryos with obvious physical damage, the oosome and pole cells could form normally (Fig. 2.6 B2, B3, white arrow head).

Of course, at a certain point, damage becomes too severe, leading to the death of the embryo. We set a stringent criterion for collecting embryos for later analysis by removing those where yolk leakage exceeded more than 10% of the embryo size (although all embryos with this amount of damage showed normal germ plasm and germline development). We also excluded embryos that showed major morphological changes as compared to uninjected embryos.
Fig. 2.6. eGFP dsRNA injected embryos as the negative control. (A) Pre-blastoderm stage. (B) Early blastoderm stage. (C) Mid-blastoderm stage. The pictures in the row of figure A1 are far-red channel, which are *in situ* hybridization results probed by *Nv-bark beetle*. The pictures in the row of figure A2 are DAPI channel (marks the nuclei). The pictures in the row of figure A3 are merged. Pictures with prime symbol are the enlarged areas marked by white boxes. The white arrow heads indicate the damage in the embryo. All the embryos are positioned with posterior side to the right and dorsal side on the top.
After removing the embryos with obvious major damage both after injections, and when performing eggshell dissection after the fixation, we were left with about 85% embryos with viable embryogenesis by the time of imaging after *in situ* hybridization. The same criteria were also applied to the experimental knockdowns where the percentages were roughly the same as in the negative control. When determining the phenotypes for the five genes, we considered the disruptions of developmental events to be potential effects of a knockdown when they were only specific to the knockdown and were never observed in the negative control. We performed the same procedures when injecting and collecting the embryos as described for the negative control. The penetrance given by the embryonic RNAi (eRNAi) knockdowns is higher than the pRNAi knockdown, ranging from 20% to 39% across experiments. Phenotypes are evident in both the pre-blastoderm and pole cell stages, for *Nv-bark*, *Nv-anillin* and *Nv-rrm*. In contrast, the oosome is not affected in pre-blastoderm embryos after *Nv-coronin* and *Nv-innexin1* dsRNA injection, with phenotypes becoming evident only after the pole cells should have formed.

**2.3.9 RNAi against three novel germ plasm components unexpectedly disrupts the oosome at from an early stage**

One transcript that particularly captured our attention was *Nv-bark*, as it was the transcript most strongly and consistently associated with PGC specification over embryonic development. However, its potential function in the germline is not clear. Since it encodes a transmembrane protein involved in epithelial junctions in *Drosophila*, we speculated that it might have a role in mediating adhesion or migration of the pole cells.
once they were formed. Surprisingly, the phenotypes produced by knocking down this transcript showed a much earlier requirement for this transcript. In early embryos (before migration of nuclei to periphery), the oosome has lost its integrity as a single unit (Fig. 2.7 A1-A3’). Instead there are scattered particles of what appears to be oosome-like material, concentrated around the lateral cortex of the embryo. Later, at the time the pole cells would normally form, no budding is observed, and some loose aggregates of germ plasm like material remains attached to the cortex (Fig. 2.6 B1-B3’). In some cases, germ plasm surrounds nuclei in a way similar to what is seen in pole cells (Fig. 2.7 C1-C3’). However, these nuclei remain part of the embryonic syncytium (Fig. 2.7 C2, C3, C2’, C3’).

Like bark, the mRNA of Drosophila anillin is not localized to the polar granules [71]. Anillin protein, however, accumulates at the base of pole cells when they are budding in Drosophia [51]. Since Anillin is a major component of the contractile ring during mitotic cytokinesis [50] and is enriched at the bud furrow during pole cell formation in Drosophila [51], we predicted that the enrichment of Nv-anillin mRNA in the oosome and pole cells would be related to Nasonia’s unique way of forming a single large pole bud instead of several small ones, as occurs in Drosophila. Surprisingly, the phenotype of Nv-anillin is indistinguishable from that of Nv-bark: The oosome does not form properly, and germ plasm material remains bound to the posterior cortex of the embryo, leading to the association of the syncytial nuclei at random posterior locations (Fig. 2.8 A1-C3’). The same set of phenotypes is observed when we knock down Nv-rrm with eRNAi (Fig. 2.9 A1-C3’). Thus, three genes with very different predicted functions all result in the same phenotype.
Fig. 2.7. Phenotypes in *Nv-bark beetle* dsRNA injected embryos. (A) Pre-blastoderm stage. (B) Early blastoderm stage. (C) Mid-blastoderm stage. The pictures in the row of figure A1 are far-red channel, which are *in situ* hybridization results probed by *Nv-nanos*. The pictures in the row of figure A2 are DAPI channel (marks the nuclei). The pictures in the row of figure A3 are merged. Pictures with prime symbol are the enlarged areas marked by white boxes. The white arrow heads indicate that the germ plasm remains part of the embryonic syncytium. All the embryos are positioned with posterior side to the right and dorsal side on the top.
It is important to note at this point that the common phenotype of the above three knockdowns is not the same as the complete loss of germ plasm activity. Such phenotypes are seen for \textit{Nv-osk}, \textit{Nv-vas}, and \textit{Nv-aubergine (Nv-aub)}. When these genes are knocked down, posterior mRNAs such as \textit{Nv-nos} take on a uniformly graded posterior cap and no enriched accumulation of germ plasm markers is ever observed at the posterior [35, 40]. This indicates that these transcripts are involved in the specific form of the oosome, but may not be essential for the production of germline-like cells.

Finally, despite the fact that the oosome does not form or migrate through the embryo as in wildtype, a bud like protrusion similar to the one that initiates pole cell formation is regularly observed (Fig. 2.8 B2-B3', Fig. 2.1 E), and this bud is not associated with the bulk of the germ plasm like material. This indicates that neither the oosome, nor its remnants, induce the bud, and that bud formation may be autonomous to the embryo. This has some precedent in \textit{Drosophila}, where the autonomous ability of the fly embryo to produce pole-cell like structures at both poles of the embryo is revealed when Arf guanine exchange factor Steppke is reduced [72]. However, in the fly, normal global repression of pole-cell formation is overcome by germ plasm components (primarily \textit{germ-cell-less}), while in \textit{Nasonia}, at least the initial budding appears to be germ plasm independent.

\textbf{2.3.10 RNAi against \textit{Nv-coronin} and \textit{Nv-innexin1} does not affect the oosome, but disrupts pole cell formation}
Fig. 2.8. Phenotypes in *Nv-anillin* dsRNA injected embryos. (A) Pre-blastoderm stage. (B) Early blastoderm stage. (C) Mid-blastoderm stage. The pictures in the row of figure A1 are far-red channel, which are *in situ* hybridization results probed by *Nv-bark beetle*. The pictures in the row of figure A2 are DAPI channel (marks the nuclei). The pictures in the row of figure A3 are merged. Pictures with prime symbol are the enlarged areas marked by white boxes. All the embryos are positioned with posterior side to the right and dorsal side on the top.
Fig. 2.9. Phenotypes in *Nv-rrm* dsRNA injected embryos. (A) Pre-blastoderm stage. (B) Early blastoderm stage. (C) Mid-blastoderm stage. The pictures in the row of figure A1 are far-red channel, which are *in situ* hybridization results probed by *Nv-bark beetle*. The pictures in the row of figure A2 are DAPI channel (marks the nuclei). The pictures in the row of figure A3 are merged. Pictures with prime symbol are the enlarged areas marked by white boxes. All the embryos are positioned with posterior side to the right and dorsal side on the top.
Fig. 2.10. Phenotypes in Nv-coronin dsRNA injected embryos. (A) Pre-blastoderm stage. (B) Early blastoderm stage. (C) Mid-blastoderm stage. The pictures in the row of figure A1 are far-red channel, which are in situ hybridization results probed by Nv-bark beetle. The pictures in the row of figure A2 are DAPI channel (marks the nuclei). The pictures in the row of figure A3 are merged. Pictures with prime symbol are the enlarged areas marked by white boxes. All the embryos are positioned with posterior side to the right and dorsal side on the top.
Fig. 2.11. Phenotypes in *Nv-innexin 1* dsRNA injected embryos. (A) Pre-blastoderm stage. (B) Early blastoderm stage. (C) Mid-blastoderm stage. The pictures in the row of figure A1 are far-red channel, which are *in situ* hybridization results probed by *Nv-bark beetle*. The pictures in the row of figure A2 are DAPI channel (marks the nuclei). The pictures in the row of figure A3 are merged. Pictures with prime symbol are the enlarged areas marked by white boxes. All the embryos are positioned with posterior side to the right and dorsal side on the top.
We tested the functions of two other transcripts with eRNAi: \textit{Nv-coronin} and \textit{Nv-innexin1}. Knockdown of both genes left the oosome intact, and able to migrate through the embryonic cytoplasm normally (Fig. 2.10 A1-A3', Fig. 2.11 A1-A3'). However, in both cases, pole cell formation fails, indicating that these genes have downstream functions that are specific to pole cell formation (Fig. 2.10 B2-B3', Fig. 2.11 B2-B3'). \textit{Nv-coronin}'s functional annotation is consistent with a role in cellularization, as it is predicted to interact with both the microtubule and actin cytoskeletons, both of which are crucial for mitosis and cell formation. While Innexin1 is most well known as a crucial component of gap junctions and a role in the nervous system [57], a role for at least one Innexin in cellularization has been demonstrated in the beetle \textit{Tribolium} [58], suggesting that the full potential of these proteins in regulating cellular processes has not been fully explored.

2.4 Discussion

2.4.1 RNA-seq analyses

Our results have uncovered an unexpectedly large amount novelty in the mRNA content of the germline determinant of the wasp \textit{Nasonia vitripennis}. This was achieved using RNAseq followed by statistical detection of differential expression of mRNAs between the anterior and posterior poles of the early embryo. The statistical predictions were then extended by \textit{in situ} hybridization and RNAi of candidate genes.

Our goal was to identify all transcripts specifically localized to the oosome, and it is worth contemplating how close we came to achieve this goal. We used three different
approaches that varied in the strategy for mapping, quantifying transcripts and assessing differential expression. These analyses agreed on the vast majority of the genes with putative significantly different enrichment at the two poles, and most of the disagreement was at the margin of differential expression that would be likely biologically significant. In addition, all analyses found genes previously identified to be localized anteriorly or posteriorly. We even detected all known genes with both strong oosome localization and strong localization at the anterior pole. We thought these could be missed, because if the anteriorly localized mRNA population approached the same levels as that in the oosome, the enrichment might be obscured.

For these reasons, we believe that we have uncovered the vast majority of mRNAs localized to the *Nasonia* oosome. Of course, no approach can guarantee comprehensiveness, and some molecules may have been missed for several reasons. For example, while the *Nasonia* genome is well annotated, it is possible that a very few genes are not represented in the genome assembly or the transcriptome annotations, and thus would not have been assessed. Indeed, we did identify slightly different sets of genes when using NCBI annotation 102 versus the OGS 2.0 annotation [49], which were created using different approaches to identify and predict genes. Most of the differences are at the margins of significance and low levels of differential expression. However, there are a handful of confirmed localized transcripts detected using annotation 102 that were missed in analyses with OGS 2.0. In addition, at least one transcript was not annotated in annotation 2.1, but was found in annotation 2.0 and was confirmed to be localized to the oosome (*Nasvi2EG001470*, Fig. 2.3 D). This is likely a false negative annotation in
annotation 102, as a very similar sequence is found in the transcriptome of the very closely related wasp *Trichomalopsis sarcophagae* [73].

Another issue that can cause false negatives is the large number of *Nasonia* transcripts that overlap, which can lead to the concatenation of transcripts. If a localized transcript is fused to a ubiquitous, highly expressed transcript, the signal of localization can be lost. This problem is largely solved by preventing novel junctions when mapping with tophat2. However, this also seem to change the calculation of significance for some transcripts, so performing the analysis with and without novel junctions gives more complete results.

Finally, there are some unknown artifacts which may cause significantly enriched genes to be missed. A prime example of this is our discovery of *Nv-coronin*. A preliminary attempt at this experiment resulted in generally poor sequencing results, and a completely unusable replicate. In general, this analysis gave predictably worse results, where several known localized genes were not found to be significant. However, all of the molecules whose functions were analyzed in this manuscript were found to be statistically significantly enriched at the posterior, including *Nv-coronin*, and we proceeded to clone them to test their localization and function. In the subsequent analyses based on high quality sequencing results (used as the basis for this manuscript), *Nv-coronin* was excluded by cuffdiff for statistical testing for unknown reasons, despite showing similar posterior enrichment that had previously been deemed statistically significant. Thus, there is a potential for false negatives if this artifact affects several genes. Very few transcripts
in the "not tested" category show a similarly strong posterior enrichment, so we believe that this effect is also small.

### 2.4.2 Comparison of the oosome of *Nasonia* to the polar granules of *Drosophila*

Germ plasm in insect embryos must perform several different functions. First, it must be able to concentrate and arrange the set of proteins and RNAs (e.g. Tud, Vas, *nos* mRNA, Piwi/Aub) that are associated with germline fate in the posterior of the egg. Second, it must ensure its own incorporation into the PGCs as they form by interacting with the specialized cytoskeletal structures that mediate pole cell formation. Germ plasm must also include molecules involved specialized features of PGCs. Such functions include the concentration and selection of mitochondria, repression of transcription, repression of transposons, and guidance of the germ cell migration to the gonad primordia.

On the other hand, we had already observed that the oosome's morphology and interaction with the cytoskeleton (during its migration and formation of the pole cells) was quite distinct from *Drosophila*, as is the way the pole cells migrate into the interior of the embryo. Therefore, we expected to find a mixture of conservation and novelty when we examined the mRNA content of the oosome. Indeed, this is what was found, but with a surprisingly strong bias toward novelty.

In terms of conservation, our analyses found *Nv-nos*, *Nv-osk* (already known factors found in both oosome and polar granules) as well as *Nv-ovo*. We also found *Nv-
The *aub* ortholog is localized as protein, but not mRNA, in *Drosophila*. While there are a large number of polar granule localized mRNAs that are not found in the oosome, we will only discuss a few that are significant.

mRNA for *polar granule component (pgc)* encodes a small peptide [74], and is strongly localized to the posterior pole [75]. Polar-granule-component protein has a crucial role in the global repression of transcription that occurs in pole cells upon their formation, through an interaction with the transcription elongation factor TEF-b [74]. This repression is a widely conserved feature of PGCs across animals, which makes it somewhat surprising that *pgc* appears to be a novelty in the *Drosophila* lineage [75].

Fascinatingly, TEF-b may be a unifying factor underlying germline quiescence, as it interacts with PIE-1 in the worm *C. elegans* to repress transcription [76]. In line with this, we have found a highly conserved transcription elongation factor, *Nv-spt5* (Fig. 3 A1-A3) localized strongly to the oosome and pole cells. In human cells, Spt5 acts as an inhibitor or transcriptional elongation, until its C-terminal domain is phosphorylated in a TEF-b dependent manner [77]. If an interaction between *Nasonia* Spt5 and TEF-b does have a role in regulating the cessation of transcription in *Nasonia* pole cells, it would be strong evidence for TEF-b being a core conserved component of the germline fate, whose interaction partners and regulators are labile across lineages.

Another crucial *Drosophila* germ cell factor that is not present in the *Nasonia* oosome is *germ cell less (gcl)*. The Germ-cell-less protein itself is very highly conserved
at the sequence level in *Nasonia*, but the mRNA showed no enrichment in our RNAseq experiments, and we independently confirmed by *in situ* hybridization that it is expressed uniformly throughout the early embryo (not shown). Gcl is important for the proper production of pole cells, apparently by regulating the orientation centrosome separation at the posterior pole, which is required for efficient pole cell formation and uptake polar granules by the pole cells [78]. At the molecular level, Gcl seems to act by downregulating torso signaling, to allow the proper conditions for pole cells to form. The lack of Gcl function in the germline of *Nasonia* is consistent with the lack of Torso signaling at the termini in the wasp [79], making the need for Gcl redundant. At the moment, it is not clear whether the use of Gcl in pole cell formation is a recent novelty in *Drosophila*, or whether it was present ancestrally, but lost in the Hymenopteran lineage.

A number of transcripts found in the *Nasonia* oosome are good candidates for generating essential PGC features. For example, a high concentration of mitochondria is a strongly conserved feature of germ plasm and PGCs across animals [80, 81]. In *Drosophila*, the long Oskar isoform plays an important role in concentrating mitochondria in the pole plasm [82]. But, since the long form of Oskar appears to be a novelty of *Drosophila* and its close relatives, other molecules should be expected to perform this role in other species. Suggestively, mRNA encoding a Milton ortholog was found strongly localized to the *Nasonia* oosome. Milton acts an adaptor that loads mitochondria onto microtubule motors for transport and localization within and between cells in *Drosophila* [83], and we propose that *Nasonia* Milton plays a role in enriching mitochondria around
the oosome and in the pole cells in the wasp, and perhaps other insect species that lack the specialization of long Oskar isoform.

Another critical function for germ cells is the control of transposable elements, which is often dependent on Tudor domain containing proteins. mRNAs for two Tudor domain proteins are present in the oosome, including \textit{Nv-qin} and \textit{Nv-tdrd7}. Neither of them is enriched in the polar granules or pole cells in \textit{Drosophila}, but both have crucial roles during oogenesis to reduce the activity of transposable elements \cite{52, 53, 84}. The presence of these additional Tudor domain encoding transcripts may indicate that either there is an increased activity of transposable elements in \textit{Nasonia} that requires an earlier response, or perhaps other mechanisms are employed in \textit{Drosophila} to combat transposon activity in the early PGCs. Further sampling of germ plasm of other insects should help to resolve these questions.

Germ cells are known to have a distinct metabolic profile from somatic cells, and this difference is related to their pluripotent stem cell-like properties, and to the requirements of their migratory properties \cite{85, 86}. Potentially related to this we have found that a \textit{Nasonia} insulin-like growth factor I mRNA is localized to the oosome. Interestingly, this mRNA encodes a short protein that shows much stronger similarity to insulin proteins of vertebrates, than it does to any of the insulin-like molecules of \textit{Drosophila} (DILPs) (JAL personal observation). In addition, an mRNA encoding a putative organic cation transporter (\textit{Nv-CG42269}) containing a Major Facilitator Superfamily (MFS) domain is strongly localized to the oosome. Such molecules are crucial for regulating
cellular metabolism and signaling at multiple levels, by controlling the trafficking of many small organic molecules (including sugars) within and between cells [87, 88]. Lipid metabolism is also uniquely regulated in germ cells, and the identification of the *Nasonia* homolog of the Acyl-CoA reductase *waterproof* may reflect this [89]. Finally, we surprisingly found a transcript encoding a protein similar to endoglucanases found in several insect lineages (but absent from Diptera). Enzymes of this type are broadly defined by their ability to break down glucose polymers. The substrate and potential role for this enzyme in the germline is not yet known.

In addition to providing insight into the conserved functions of germ cell components, we also found several molecules that do not have clear homology outside of the Hymenoptera, or in some cases outside of *Nasonia* and its closest relatives. This includes a novel RNA recognition domain containing protein whose function we analyzed in depth (*Nv-rrm* discussed in the following section). We also found that an mRNA encoding an ankyrin domain protein (*Nv-OoCLANK (LOC100679945)*) that belongs to a family of proteins that underwent a massive amplification within chalcid wasp lineage, and which appears to have entered the ancestral chalcid wasp by horizontal transfer [43]. Finally, a handful of transcripts have no identifiable domains or homologs. The functional relevance of these molecules will be an area of intense interest in the future.

**2.4.3 Unexpected functions of novel oosome components**
Early in our analysis we chose a handful of transcripts for functional analysis. These were chosen based on a combination of criteria that included high enrichment in the oosome, novelty, and the potential to give phenotypes that we could characterize with the current set of functional tools available to us in *Nasonia*. Three of these (*Nv-bark, Nv-anillin* and *Nv-rrm*) gave an unexpected phenotype, where oosome-like material was not coalesced into the typical spherical oosome structure, but rather was scattered in clumps attached to the plasma membrane near the posterior pole. Eventually, these clumps of germ plasm-like material come into contact with syncytial nuclei when they migrate to the cortex. However, this remnant material is unable to induce the pole cell fate. It is important to reiterate that this phenotype is quite distinct from that seen for genes that have a core role in oosome assembly (*Nv-osk, Nv-tud, Nv-vas* and *Nv-aub* (not shown). In these cases, there is no hint of the oosome, and mRNAs normally localized in the oosome are distributed in homogenous caps at the posterior pole, rather than as discrete clumps of material [35].

These phenotypes indicate these genes are involved in the coalescence of the oosome into single entity within the central column of cytoplasm in the embryo, and/or maintenance of oosome integrity. The molecular bases of such functions are not completely clear at the moment for any of these three genes. *Nasonia* Bark is a putative transmembrane protein and would be predicted to be targeted to the membrane. One potential hypothesis is that *Nasonia* Bark is targeted to the membrane at the posterior pole of the embryo, releases and/or repels oosome material from adhering to the plasma membrane, forcing the oosome to remain in the bulk cytoplasm where it concentrated into
a large sphere and moved around by strong cytoplasmic flows that occur during the earliest cleavages. This model would also imply that interaction with the cortex prevents oosome material from coalescing, leading to the scattered clumps we observe in Nv-bark knockdowns. Mechanistically, this could be related to the ability of Bark to induce endocytosis [90], a process that is associated with proper anchoring of the germ plasm and the recruitment of specialized actin binding proteins to the posterior pole in Drosophila [91].

An alternative hypothesis is that Nasonia Bark produced in the oosome is not secreted but is instead incorporated as an important structural component of the large, solid form of the oosome. This is consistent with the structure of the protein, which contains protein-protein interaction domains of different types [45, 46]. Testing these hypotheses will require in depth analysis of the subcellular localization of Nasonia Bark during early embryogenesis, and proteomic analysis of binding partners of Nasonia Bark.

The role of Nv-anillin in maintaining the stability of the oosome was also surprising and the molecular basis of the phenotype will require further investigation. Anillin orthologs are well known as actin binding proteins involved in assembling the contractile ring required to separate cells in cytokinesis [50]. Anillin also plays a crucial and novel role in the specialized cytokinesis of the Drosophila pole cells [51]. While these known functions might have indicated that oosome localization of Nv-anillin was related to an important role in the specialized polar bud formed in Nasonia, RNAi showed that this protein has an earlier role in oosome assembly/maintenance (Fig. 2.8 A1-A3'). Similar to
Nasonia Bark, one possible function of Nasonia Anillin is as a structural component of the oosome, which may or may not be related to its ability to bind actin and associated proteins. Alternatively, Nv-anillin may act to release and/or repel the oosome from the cortex, as proposed above for Nv-bark.

Anillin homologs have known functions which are directly related to the processes of release of germ plasm from the embryonic plasma membrane. In Drosophila, the germ plasm is tightly bound to the plasma membrane until the nuclei reach the posterior pole of the embryo. The centrosomes associated with these nuclei mediate detachment of the pole plasm from the cortex through interactions of the astral microtubules emanating from the centrosomes [92]. This is in contrast to Nasonia, where the oosome detaches from the cortex at about the same time as the zygotic nucleus begins its first division at the anterior pole. While nuclei are lacking at this time, numerous centrosomes are present, as they are provided maternally in a process characteristic of many Hymenopteran embryos [93]. Nasonia Anillin could be relevant to a model where astral microtubules emanating from maternally provided centrosomes detach oosome material from the cortex, because Anillin homologs have been shown to mediate interactions between the actin cytoskeleton and cortical and subcortical microtubule arrays in multiple model systems [94, 95]. Again, in depth examination of the subcellular localization and interactions of Nasonia Anillin will be required to completely understand its role in maintaining the oosome.
Based on the presence of only RNA recognition motifs in the protein, we predict that \textit{Nv-rrm} will have one of two likely roles. One possibility is that it is involved in translational regulation of key regulators of oosome structure, presumably including \textit{Nasonia} Anillin and \textit{Nasonia} Bark. Alternatively, \textit{Nv-rrm} may be important in binding RNA and protein in order to maintain the structural integrity of the oosome. In addition, neither of these possibilities is mutually exclusive.

The knockdowns of \textit{Nv-coronin} and \textit{Nv-innexin1} had specific effects only on the formation of pole cells, while the oosome appeared to remain intact. Coronin is an actin binding protein associated with the formation of highly concentrated networks of F-actin \cite{65}. It seems likely that \textit{Nasonia} Coronin has an important role in organizing an actin cytoskeleton arrangement specialized for the formation of the large polar bud that initiates pole cell formation.

The potential role of \textit{Nv-innexin1} is somewhat more mysterious. Innexins are typically known as components of the gap junctions that are found in some tightly integrated epithelial tissues. Such junctions would not be expected of the motile pole cells. Interestingly, Innexin-7 (a paralogous protein with a similar structure to \textit{Nasonia} Innexin1) in the beetle \textit{Tribolium} has a novel role in cellularization of the syncytial blastoderm. Such a function for \textit{Nasonia} Innexin1 would explain the failure in pole cell formation we see after RNAi.
2.5 Conclusion

This work has revealed numerous unexpected mRNAs that are localized to the germ plasm of the wasp *Nasonia*. The results have given insights on the potentially ancestral mechanisms used by germ plasm to accomplish conserved required functions, such as a possible ancestral role for Milton orthologs in bringing mitochondria to the germ plasm, a function replaced by the long Oskar isoform in *Drosophila*. On the other hand, our results have identified numerous components that are likely to be specific to *Nasonia* and its relatives in the parasitic wasp lineage (e.g., *Nv-OoCLANK*, *Nv-rrm*, and the use of *Nv-Bark* and *Nv-Anillin* in assembling the oosome). Deeper analyses of the functions of these molecules in *Nasonia*, and broader sampling of germ plasm in other holometabolous insects will be required to determine the patterns of evolutionary change in the germ plasm. Such analyses will be important because the germ plasm is a uniquely powerful organelle that can rapidly drive naive nuclei into a highly specialized, yet functionally totipotent state. Understanding how and why such a fundamental substance changes and is even lost in the course of evolution will provide foundational insights into the mechanisms of cell fate determination and the interaction of subcellular organelles and their cellular milieu.

2.6 Methods

2.6.1 Sample preparation
In order to identify the components of the maternally deposited mRNAs in the oosome located in the posterior half of the parasitoid wasp *Nasonia vitripennis* embryos, we collected and bisected the pre-blastoderm stage embryos (0-2 hours after egg lay at 25 °C) to detect the differential expression levels between the anterior and posterior halves of the embryos.

The embryos were aligned with the anterior pole to the right on the ice-prechilled and heptane glue-coated slide. Then a thin layer of halocarbon oil 700 (Sigma) was applied to cover the embryos. The slide was transferred on the dry ice-prechilled “guillotine” and then was anchored by tightening the screws on each end. After the halocarbon oil 700 was solidified on dry ice, put the guillotine on the dry ice-prechilled stabilizer and transferred it under the dissection microscope. The embryos were positioned to match the slot in the guillotine where the dry ice-prechilled razor blade will be inserted into. After the embryos were cut, the anterior and posterior halves of the embryos were immediately collected and transferred into the two 1.5mL non-stick RNase-Free microfuge tubes (Ambion) with the dry ice-prechilled probe, separately. Three biological replicates were created.

Total RNA was isolated from these six samples for library preparation. In the library preparation upon which this manuscript is based, we used around one microgram of total RNA from each sample, in which 100ng of total RNA was from the *Nasonia* other 900ug was from a distantly related parasitic wasp (*Melittobia digitata* [43]). Libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB
Libraries were validated and quantified before being pooled and sequenced on an Illumina HiSeq 2000 sequencer with a 100 bp paired-end protocol. Sequence files are available in the NCBI SRA database under accession SRP156232.

2.6.2 RNA sequencing data analysis

The quality of the sequencing data was determined using FastQC software. The sequences were processed by Cufflinks package for differential expression detection, using multiple variations on the default parameters (job files in Supplemental file 2.7.2). Briefly, raw reads were aligned to annotation 102 of assembly 2.1 of the *N. vitripennis* genome(https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Nasonia_vitripennis/102/) using TopHat2[96]. These results were either used directly in cuffdiff [97], or were further processed using stringtie [98] to generate new transcriptome predictions and quantification. Various normalization parameters were used, and each permutation gave slightly different results. In addition, we mapped the reads using assembly 1.0 and annotation 2.0 (OGS 2.0) [49]. All of the commands jobs in this process of analysis are included in the Supplemental file 2.7.2. The computing work was done by the High Performance Computing Cluster located at University of Illinois at Chicago.

Probes and dsRNAs for the chosen genes were generated by the protocol described in [40]. Primers for generating these templates are provided in Supplemental
file 2.7.3. Alkaline phosphatase *in situ* hybridization was performed by the protocol described in [99].

**2.6.2 Embryonic RNA interference (eRNAi)**

In order to perform eRNAi on the early *Nasonia* embryos to study the germline candidate genes’ functions, we created the following workflow:

Around 30 pre-blastoderm stage embryos (0-1 hour after egg lay at 25 °C) were collected and quickly aligned vertically on the heptane glue-coated 18mm×18mm coverslip. This coverslip was transferred and anchored on the ice-prechilled slide by applying a thin layer of water. The slide was then put in an air tight petri dish with proper amount of desiccant (Drierite with indicator, 8 mesh, ACROS Organics) pre-chilled at 4 °C. The embryos were dehydrated in the desiccant at 4 °C for 45 minutes. After dehydration, the embryos were covered with a layer of halocarbon oil 700 and were ready for microinjection.

The dsRNAs were dissolved in Nuclease-Free Water (Ambion) at the concentration of 1mg/mL and loaded into the Femtotips II Microinjection Capillary (Cat. No. 930000043, Eppendorf). The constant pressure was set at 500hpa and the injection pressure was set initially at 250hpa with periodic adjustment as the needle changed over the course of injection. The process of injection was performed at room temperature and needed to be done as soon as possible. After injection, the slide was transferred into a
paper towel-moisturized petri dish pre-warmed at 28 °C to incubate the injected embryos for specific developmental stages. The embryogenesis of these embryos was stopped at pre-blastoderm stage (before the budding), beginning of blastoderm stage (during budding), and later in blastoderm stage (pole cells formed and/or after pole cell divisions). To stop the development, the coverslip was put into the heptane to wash off the halocarbon oil 700 for three minutes, and then transferred into the 37% formaldehyde-saturated heptane for 2-5 hours fixation in the dark with the embryos facing up.

After fixation, the coverslip was carefully taken out of the fixative, and flipped upside down to gently press the embryos on a double-sided tape that was taped on a petri dish, so that all the embryos can be anchored on the tape for dissection. Add about 15mL PBS with 1% Tween, use the needle (BD PrecisionGlide Needle, 30G × 1) to carefully remove the eggshells from the embryos. The dissected embryos were then transferred by pipette into the 1.5mL non-stick RNase-Free microfuge tubes. The embryos were immediately dehydrated by 100% Methanol and stored at -20 °C.

Before performing fluorescent in situ hybridization (FISH) on those eRNAi knocked out embryos, they need to be rehydrated with a series of Methanol/PBT washes (75%, 50%, 25%). The protocol for FISH was adapted from [40]. A detailed protocol is available on request.

2.7 Supplementary materials
2.7.1 Significantly enriched transcripts from both experiments.

The transcripts listed in “Compilation of Transcripts significant in the main analyses” were generated using the sequencing data from the second experiment by NCBI annotation 102 with various normalization parameters. The “Transcripts from the second experiment only found in OGS 2.0” were the transcripts from the second experiment only found using the annotation OGS 2.0, but not found using NCBI annotation 102. The “Transcripts from the first experiment” were the transcripts only found in the first time sequencing data, but not in the second time sequencing data.

<table>
<thead>
<tr>
<th>NCBI annotation 102 name</th>
<th>Name applied in manuscript</th>
<th>OGS 2.0 name</th>
<th>( \text{log}_{2}(\text{fold change}) )</th>
<th>( p ) value</th>
<th>( q ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC100120906</td>
<td>not tested</td>
<td>Nasvi2EG001444</td>
<td>-3.22271</td>
<td>5.00E-05</td>
<td>0.00430</td>
</tr>
<tr>
<td>LOC100680515</td>
<td></td>
<td>Nasvi2EG022897</td>
<td>-3.20827</td>
<td>5.00E-05</td>
<td>0.00430</td>
</tr>
<tr>
<td>LOC100102554</td>
<td>giant</td>
<td>Nasvi2EG000388</td>
<td>-2.84997</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100121206</td>
<td>mex3</td>
<td>Nasvi2EG006980</td>
<td>-2.59922</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100114679</td>
<td>unsuccessful cloning</td>
<td>Nasvi2EG017534</td>
<td>-2.55947</td>
<td>5.00E-05</td>
<td>0.00430</td>
</tr>
<tr>
<td>LOC100119882</td>
<td>very strong anterior early, then gone at blastoderm</td>
<td>Nasvi2EG007110</td>
<td>-2.29605</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC107982211</td>
<td></td>
<td>Nasvi2EG007111</td>
<td>-2.29605</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100123633</td>
<td>unsuccessful cloning</td>
<td>Nasvi2EG003943</td>
<td>-1.92542</td>
<td>5.00E-05</td>
<td>0.00430</td>
</tr>
<tr>
<td>LOC100678694</td>
<td>not tested</td>
<td>Nasvi2EG008558</td>
<td>-1.91378</td>
<td>5.00E-05</td>
<td>0.00434</td>
</tr>
<tr>
<td>LOC100122183, (LOC100122201)</td>
<td>83 not tested, LOC1001222 01 ubiquitous in preblastoderm</td>
<td>Nasvi2EG008559, (Nasvi2EG008551, Nasvi2EG008563)</td>
<td>-1.73437</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100121288</td>
<td>expressed in the anterior pole</td>
<td>Nasvi2EG003644</td>
<td>-1.57319</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100120121</td>
<td>Exd2, anterior</td>
<td>Nasvi2EG011856</td>
<td>-1.5326</td>
<td>5.00E-05</td>
<td>0.00434</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
<td>Log2 Fold Change</td>
<td>P-value</td>
<td>False Discovery Rate</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------------</td>
<td>------------------</td>
<td>---------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td>LOC100313502</td>
<td>localized very strong Nasvi2EG002890</td>
<td>-1.47888</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td></td>
</tr>
<tr>
<td>LOC100115727</td>
<td>not expressed in pre-blastoderm stage, expressed in the pattern of rings in blastoderm stage Nasvi2EG028970</td>
<td>-1.38593</td>
<td>5.00E-05</td>
<td>0.00430</td>
<td></td>
</tr>
<tr>
<td>LOC100121774</td>
<td>no specific expression Nasvi2EG013933</td>
<td>-1.3142</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td></td>
</tr>
<tr>
<td>LOC103315431</td>
<td>Para, not tested Nasvi2EG022279</td>
<td>-1.15289</td>
<td>5.00E-05</td>
<td>0.00434</td>
<td></td>
</tr>
<tr>
<td>LOC100101930</td>
<td>ubiquitous and strong expression in preblastoderm stage Nasvi2EG013461</td>
<td>-1.05247</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td></td>
</tr>
<tr>
<td>LOC100118628</td>
<td>unsuccessful cloning Nasvi2EG004048</td>
<td>-0.938123</td>
<td>0.00</td>
<td>0.02984</td>
<td></td>
</tr>
<tr>
<td>LOC100121300</td>
<td>not tested Nasvi2EG004082, Nasvi2EG004083, Nasvi2EG004084</td>
<td>-0.909109</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td></td>
</tr>
<tr>
<td>LOC100123508</td>
<td>anterior localization early, DV and neuroblasts later Nasvi2EG013803</td>
<td>-0.95121</td>
<td>5.00E-05</td>
<td>0.00434</td>
<td></td>
</tr>
<tr>
<td>LOC100123856</td>
<td>not tested Nasvi2EG002296</td>
<td>-0.900637</td>
<td>0.00</td>
<td>0.01981</td>
<td></td>
</tr>
<tr>
<td>LOC100121321</td>
<td>not tested Nasvi2EG006339, Nasvi2EG006340</td>
<td>-0.860341</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td></td>
</tr>
<tr>
<td>LOC100118491</td>
<td>slowpoke, no specific expression weakly expressed in anterior Nasvi2EG007782</td>
<td>-0.849958</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td></td>
</tr>
<tr>
<td>LOC100119011</td>
<td>no specific localization Nasvi2EG011242</td>
<td>-0.763556</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td></td>
</tr>
<tr>
<td>LOC100122558</td>
<td>not tested Nasvi2EG002984</td>
<td>-0.748257</td>
<td>5.00E-05</td>
<td>0.00434</td>
<td></td>
</tr>
<tr>
<td>LOC100115322</td>
<td>not tested Nasvi2EG002984</td>
<td>-0.749067</td>
<td>5.00E-05</td>
<td>0.00434</td>
<td></td>
</tr>
<tr>
<td>LOC100115729</td>
<td>Ubiquitously expressed in pre-blastoderm stage, enriched at anterior pole and weakly Nasvi2EG013769</td>
<td>-0.718655</td>
<td>5.00E-05</td>
<td>0.00430</td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
<td>Nasvi2EG000804</td>
<td>LOC100124035</td>
<td>LOC100120674</td>
<td>LOC100117527</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------</td>
<td>----------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>LOC100124035</td>
<td>expressed in the oosome</td>
<td>Nasvi2EG000804</td>
<td>-0.694832</td>
<td>0.000</td>
<td>0.01432</td>
</tr>
<tr>
<td>LOC100120674</td>
<td>Nasvi2EG009411</td>
<td>-0.690227</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td>494</td>
</tr>
<tr>
<td>LOC100117527</td>
<td>unsuccessful cloning</td>
<td>Nasvi2EG001295</td>
<td>-0.687162</td>
<td>0.000</td>
<td>0.02961</td>
</tr>
<tr>
<td>LOC100116205</td>
<td>not tested</td>
<td>-0.662914</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td>494</td>
</tr>
<tr>
<td>LOC100119625</td>
<td>not tested</td>
<td>-0.65155</td>
<td>0.001</td>
<td>0.04995</td>
<td>94</td>
</tr>
<tr>
<td>LOC100121045</td>
<td>not tested</td>
<td>-0.625302</td>
<td>0.000</td>
<td>0.03791</td>
<td>48</td>
</tr>
<tr>
<td>LOC103317367</td>
<td>not tested</td>
<td>-0.607705</td>
<td>0.000</td>
<td>0.03939</td>
<td>494</td>
</tr>
<tr>
<td>LOC100117144</td>
<td>not tested</td>
<td>-0.592524</td>
<td>0.000</td>
<td>0.04554</td>
<td>53</td>
</tr>
<tr>
<td>LOC100120131</td>
<td>not tested</td>
<td>-0.586312</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td>494</td>
</tr>
<tr>
<td>LOC100120439</td>
<td>not tested</td>
<td>-0.570438</td>
<td>5.00E-05</td>
<td>0.00434</td>
<td>81</td>
</tr>
<tr>
<td>LOC100679839</td>
<td>not tested</td>
<td>-0.549581</td>
<td>5.00E-05</td>
<td>0.00434</td>
<td>81</td>
</tr>
<tr>
<td>LOC100123255</td>
<td>not tested</td>
<td>-0.548008</td>
<td>5.00E-05</td>
<td>0.00430</td>
<td>633</td>
</tr>
<tr>
<td>LOC100115388</td>
<td>Tm1</td>
<td>-0.543346</td>
<td>0.000</td>
<td>0.03955</td>
<td>81</td>
</tr>
<tr>
<td>LOC100118986</td>
<td>not tested</td>
<td>-0.540171</td>
<td>0.000</td>
<td>0.02864</td>
<td>88</td>
</tr>
<tr>
<td>LOC100678582</td>
<td>not tested</td>
<td>-0.52989</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td>494</td>
</tr>
<tr>
<td>LOC10012002</td>
<td>not tested</td>
<td>-0.526909</td>
<td>5.00E-05</td>
<td>0.00434</td>
<td>81</td>
</tr>
<tr>
<td>LOC100116851</td>
<td>not tested</td>
<td>-0.52675</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td>494</td>
</tr>
<tr>
<td>LOC100119465</td>
<td>not tested</td>
<td>-0.525578</td>
<td>5.00E-05</td>
<td>0.00434</td>
<td>81</td>
</tr>
<tr>
<td>LOC100678441</td>
<td>not tested</td>
<td>-0.517297</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td>494</td>
</tr>
<tr>
<td>LOC100121959</td>
<td>not tested</td>
<td>-0.510554</td>
<td>0.000</td>
<td>0.01120</td>
<td>11</td>
</tr>
<tr>
<td>LOC100678448</td>
<td>not tested</td>
<td>-0.50581</td>
<td>0.000</td>
<td>0.00789</td>
<td>655</td>
</tr>
<tr>
<td>LOC100118206</td>
<td>not tested</td>
<td>-0.50551</td>
<td>0.000</td>
<td>0.00737</td>
<td>553</td>
</tr>
<tr>
<td>LOC100123602</td>
<td>not tested</td>
<td>-0.502997</td>
<td>5.00E-05</td>
<td>0.00389</td>
<td>494</td>
</tr>
<tr>
<td>LOC100121846</td>
<td>not tested</td>
<td>-0.4932</td>
<td>0.000</td>
<td>0.04315</td>
<td>97</td>
</tr>
<tr>
<td>LOC100122143</td>
<td>CPR39</td>
<td>Nasvi2EG010007</td>
<td>-0.492529</td>
<td>0.000</td>
<td>0.02370</td>
</tr>
<tr>
<td>LOC1001235287</td>
<td>not tested</td>
<td>-0.485824</td>
<td>0.000</td>
<td>0.02621</td>
<td>72</td>
</tr>
<tr>
<td>LOC100121267</td>
<td>not tested</td>
<td>-0.485033</td>
<td>0.000</td>
<td>0.01873</td>
<td>78</td>
</tr>
<tr>
<td>LOC100121632</td>
<td>not tested</td>
<td>-0.476594</td>
<td>0.000</td>
<td>0.01604</td>
<td>86</td>
</tr>
<tr>
<td>LOC100123602</td>
<td>not tested</td>
<td>-0.471264</td>
<td>0.000</td>
<td>0.03939</td>
<td>52</td>
</tr>
<tr>
<td>LOC100120395</td>
<td>not tested</td>
<td>-0.46058</td>
<td>0.000</td>
<td>0.03327</td>
<td>84</td>
</tr>
<tr>
<td>LOC100121098</td>
<td>not tested</td>
<td>-0.460201</td>
<td>0.000</td>
<td>0.03202</td>
<td>55</td>
</tr>
<tr>
<td>LOC100121632</td>
<td>not tested</td>
<td>-0.458016</td>
<td>0.000</td>
<td>0.03118</td>
<td>5</td>
</tr>
<tr>
<td>LOC100120252</td>
<td>not tested</td>
<td>-0.445218</td>
<td>0.000</td>
<td>0.01039</td>
<td>95</td>
</tr>
<tr>
<td>LOC100122962</td>
<td>not tested</td>
<td>-0.444901</td>
<td>0.000</td>
<td>0.01416</td>
<td>49</td>
</tr>
<tr>
<td>LOC</td>
<td>Status</td>
<td>Gene ID</td>
<td>Expression Level</td>
<td>P-value</td>
<td>Q-value</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>LOC100121417</td>
<td>not tested</td>
<td>Nasvi2EG001615</td>
<td>0.442791</td>
<td>0.000</td>
<td>0.0458</td>
</tr>
<tr>
<td>LOC100120759</td>
<td>not tested</td>
<td>Nasvi2EG000502</td>
<td>0.448054</td>
<td>0.000</td>
<td>0.03118</td>
</tr>
<tr>
<td>LOC100119315</td>
<td>not tested</td>
<td>Nasvi2EG0003287</td>
<td>0.471371</td>
<td>0.000</td>
<td>0.0125</td>
</tr>
<tr>
<td>LOC100121362</td>
<td>not tested</td>
<td>Nasvi2EG0005420</td>
<td>0.516264</td>
<td>5.00E-05</td>
<td>0.00430</td>
</tr>
<tr>
<td>LOC100121878</td>
<td>not tested</td>
<td>Nasvi2EG013849</td>
<td>0.5263</td>
<td>0.000</td>
<td>0.01981</td>
</tr>
<tr>
<td>LOC100123551</td>
<td>weak in oosome</td>
<td>Nasvi2EG000641</td>
<td>0.511357</td>
<td>0.000</td>
<td>0.03595</td>
</tr>
<tr>
<td>LOC100118283</td>
<td>not tested</td>
<td>Nasvi2EG003652</td>
<td>0.511896</td>
<td>0.000</td>
<td>0.0378</td>
</tr>
<tr>
<td>LOC100119315</td>
<td>no specific localization</td>
<td>Nasvi2EG013964</td>
<td>0.490021</td>
<td>0.000</td>
<td>0.01039</td>
</tr>
<tr>
<td>LOC100121362</td>
<td>not tested</td>
<td>Nasvi2EG007644</td>
<td>0.512927</td>
<td>0.000</td>
<td>0.00789</td>
</tr>
<tr>
<td>LOC100115356</td>
<td>not tested</td>
<td>Nasvi2EG013849</td>
<td>0.526525</td>
<td>5.00E-05</td>
<td>0.00434</td>
</tr>
</tbody>
</table>

**Descriptions:**
- **LOC100121417**: not tested
- **LOC100120759**: not tested
- **LOC100114310**: not tested
- **LOC100119315**: no specific localization
- **LOC100123551**: weak in oosome
- **LOC100118283**: not tested
- **LOC100121362**: not tested
- **LOC100118272**: ubiquitous and strong expression in preblastoderm stage
- **LOC100118715**: ubiquitous in preblastoderm stage
- **LOC100123811**: ubiquitous and strong expression in preblastoderm stage
- **LOC100121878**: not tested
- **LOC100115356**: not tested
<p>| LOC100114688 | strong and ubiquitous expression in pre-blastoderm stage | Nasvi2EG006689 | 0.544467 [5.00E-05] | 0.00389 | 494 |
| LOC100115630 | no specific localization | Nasvi2EG026339 | 0.545728 | 0.000 | 0.02709 |
| LOC100678185 | not tested | Nasvi2EG000799 | 0.564636 | 0.000 | 0.01121 |
| LOC100124030 | weakly enriched in posterior, weak and transient expression in oosme | Nasvi2EG000881 | 0.568069 | 5.00E-05 | 0.00389 |
| LOC103317596 | not tested | Nasvi2EG005622 | 0.584679 | 5.00E-05 | 0.00389 |
| LOC100116063 | no specific localization | Nasvi2EG011215 | 0.587278 | 0.000 | 0.00791 |
| LOC100679158 | ubiquitously expressed in pre-blastoderm stage with elevated expression in the posterior pole | Nasvi2EG018255 | 0.592265 | 0.000 | 0.03847 |
| LOC100119585 | ubiquitously expressed in pre-blastoderm stage | Nasvi2EG012736 | 0.611428 | 5.00E-05 | 0.00434 |
| LOC100679126 | not tested | Nasvi2EG008592 | 0.619314 | 0.000 | 0.02451 |
| LOC100121645,LOC100680502 | weak and ubiquitous expression in pre-blastoderm stage with elevated expression in the ososome | Nasvi2EG008505,Nasvi2EG008506 | 0.627951 | 5.00E-05 | 0.00737 |
| LOC1000677870 | not tested | Nasvi2EG006679 | 0.636612 | 5.00E-05 | 0.00389 |
| LOC100121779 | not tested | Nasvi2EG007200 | 0.654268 | 0.000 | 0.03118 |
| LOC100120993 | not tested | Nasvi2EG006803 | 0.680119 | 0.000 | 0.01120 |
| LOC100117750 | weakly expressed in posterior pole, not specific in ososome during preblastoderm, not in other stages | Nasvi2EG004772,Nasvi2EG004773 | 0.687087 | 5.00E-05 | 0.00434 |
| LOC100118414 | aubergine | Nasvi2EG0007365 | 0.688636 | 5.00E-05 | 0.00430 |
| LOC100119304 | no specific localization | Nasvi2EG014512 | 0.690883 | 0.000 | 0.01120 |</p>
<table>
<thead>
<tr>
<th>GenBank AC</th>
<th>Description</th>
<th>Log2FoldChange</th>
<th>P-Value</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC100121187</td>
<td>Nv-CG4552 Nasvi2EG009916</td>
<td>0.70076</td>
<td>0.000</td>
<td>0.03680</td>
</tr>
<tr>
<td>LOC100680197</td>
<td>NvEG005246</td>
<td>0.723066</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100114138</td>
<td>not tested Nasvi2EG031989</td>
<td>0.73701</td>
<td>0.000</td>
<td>0.04315</td>
</tr>
<tr>
<td>LOC100122778</td>
<td>weakly expressed in posterior pole, not specific in oosme during preblastoder m, not in other stages Nasvi2EG005059</td>
<td>0.775559</td>
<td>5.00E-05</td>
<td>0.00434</td>
</tr>
<tr>
<td>LOC100118351</td>
<td>not tested Nasvi2EG022625</td>
<td>0.788854</td>
<td>0.000</td>
<td>0.04315</td>
</tr>
<tr>
<td>LOC100118857</td>
<td>waterproof Nasvi2EG015690</td>
<td>0.801352</td>
<td>5.00E-05</td>
<td>0.00430</td>
</tr>
<tr>
<td>LOC100113954</td>
<td>weakly expressed in posterior pole, not specific in oosme during preblastoder m. Strong and ubiquitous expression in blastoderm stage, not expressed in the pole cells Nasvi2EG027754, Nasvi2EG027756</td>
<td>0.834995</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100115154</td>
<td>no specific localization old1 Nasvi2EG013231</td>
<td>0.850862</td>
<td>5.00E-05</td>
<td>0.00434</td>
</tr>
<tr>
<td>LOC100107367</td>
<td>N/A Nasvi2EG014289</td>
<td>0.882696</td>
<td>0.000</td>
<td>0.00791</td>
</tr>
<tr>
<td>LOC100117772</td>
<td>weak in oosme zen Nasvi2EG000058</td>
<td>0.914907</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100117357</td>
<td>not tested Nasvi2EG020485,Nasvi2EG020490</td>
<td>0.927156</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100121335</td>
<td>Nasvi2EG009307, Nasvi2EG009308 unsuccessful cloning anillin Nasvi2EG006868</td>
<td>0.975877</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100119116</td>
<td>Nv-CG42269 Nasvi2EG010895</td>
<td>1.1881</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100677932</td>
<td>Nasvi2EG016466</td>
<td>1.00765</td>
<td>5.00E-05</td>
<td>0.00430</td>
</tr>
<tr>
<td>LOC100679945</td>
<td>Nv-OoCLANK Nasvi2EG016126</td>
<td>1.02046</td>
<td>5.00E-05</td>
<td>0.00434</td>
</tr>
<tr>
<td>LOC100121226</td>
<td>tdrd7 Nasvi2EG005860</td>
<td>1.07058</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100119940</td>
<td>Nv-CG2201 Nasvi2EG016344</td>
<td>1.08918</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>LOC100115818</td>
<td>spt5 Nasvi2EG018250</td>
<td>1.0983</td>
<td>0.000</td>
<td>0.04315</td>
</tr>
<tr>
<td>LOC100123906</td>
<td>dpp Nasvi2EG000703</td>
<td>1.13743</td>
<td>0.000</td>
<td>0.04502</td>
</tr>
<tr>
<td>LOC100120350</td>
<td>Nv-CG42269 Nasvi2EG010895</td>
<td>1.1881</td>
<td>5.00E-05</td>
<td>0.00389</td>
</tr>
<tr>
<td>GenBank Accession</td>
<td>Gene Name</td>
<td>Description</td>
<td>Log2 Fold Change</td>
<td>P-Value</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>LOC100120940</td>
<td>mp1</td>
<td>Nasvi2EG006450</td>
<td>1.29315</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100122980</td>
<td>elbow</td>
<td>Nasvi2EG005511</td>
<td>1.38151</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100680502</td>
<td>weakly expressed in the posterior pole, not specific in oosome</td>
<td>Nasvi2EG008506</td>
<td>1.44713</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100120333</td>
<td>innexin1</td>
<td>Nasvi2EG015809</td>
<td>1.5446</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100679703</td>
<td>insulin-like growth factor I (igf)</td>
<td>Nasvi2EG009280</td>
<td>1.57675</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100678249</td>
<td>qin</td>
<td>Nasvi2EG016574</td>
<td>1.64127</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC1007982088</td>
<td>not tested</td>
<td>Nasvi2EG016575</td>
<td>1.64127</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100680462</td>
<td>Not expressed in pre-blastoderm stage, expressed in both poles in blastoderm stage</td>
<td>Nasvi2EG007407</td>
<td>1.65009</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100678440</td>
<td>rm</td>
<td>Nasvi2EG019633</td>
<td>1.73465</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100680147</td>
<td>very weak expression in the posterior pole and the oosome</td>
<td>Nasvi2EG003874</td>
<td>1.80967</td>
<td>0.0002</td>
</tr>
<tr>
<td>LOC100113617</td>
<td>C16orf87</td>
<td>Nasvi2EG009070</td>
<td>1.86969</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100116315</td>
<td>Caudal</td>
<td>Nasvi2EG037157</td>
<td>2.28481</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100679293</td>
<td>ovo, weak and transient expression in oosome, posterior expression in pre-blastoderm stage. Ring expression pattern and dv pattern in blastoderm stage</td>
<td>Nasvi2EG009257</td>
<td>3.05441</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100188942</td>
<td>not tested</td>
<td>Nasvi2EG003556</td>
<td>3.29953</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100120457</td>
<td>not tested</td>
<td>Nasvi2EG013570</td>
<td>4.21801</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100678574</td>
<td>not tested</td>
<td>Nasvi2EG013954</td>
<td>4.30037</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100500744</td>
<td>oskar</td>
<td>Nasvi2EG013571</td>
<td>4.32224</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100120269</td>
<td>bark beetle (bark)</td>
<td>Nasvi2EG003216</td>
<td>5.59301</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100104006</td>
<td>nanos</td>
<td>Nasvi2EG022188</td>
<td>5.78008</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>LOC100678931</td>
<td>likely artifact</td>
<td>Nasvi2EG021838</td>
<td>#NAME?</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Total: 159 transcripts
### Transcripts from the second experiment only found in OGS 2.0

<table>
<thead>
<tr>
<th>NCBI annotation</th>
<th>Name applied in manuscript</th>
<th>OGS 2.0 name</th>
<th>log2(fold_change)</th>
<th>p_value</th>
<th>q_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>variable expressions</td>
<td>Nasvi2EG029760</td>
<td>-1.45513</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>N/A</td>
<td>unsuccessful cloning</td>
<td>Nasvi2EG010608</td>
<td>-2.08639</td>
<td>0.0005</td>
<td>0.0372967</td>
</tr>
<tr>
<td>N/A</td>
<td>no specific localization</td>
<td>Nasvi2EG008552</td>
<td>-1.753</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>N/A</td>
<td>ubiquitous expression in pre-blastoderm stage unsuccessful cloning</td>
<td>Nasvi2EG008556</td>
<td>-1.753</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>N/A</td>
<td>no specific localization</td>
<td>Nasvi2EG029763</td>
<td>-1.46663</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>LOC100677844</td>
<td>no specific localization</td>
<td>Nasvi2EG002891</td>
<td>-1.1966</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>N/A</td>
<td>weakly expressed in anterior pole in pre-blastoderm stage unsuccessful cloning</td>
<td>Nasvi2EG011217</td>
<td>-1.08831</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>LOC107980511</td>
<td>Ubiquitously and anteriorly expressed in pre-blastoderm stage unsuccessful cloning</td>
<td>Nasvi2EG016180</td>
<td>-1.05808</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>LOC100120099</td>
<td>Ubiquitously and anteriorly expressed in pre-blastoderm stage unsuccessful cloning</td>
<td>Nasvi2EG016181</td>
<td>-1.05808</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>LOC100122199</td>
<td>Ubiquitously and anteriorly expressed in pre-blastoderm stage unsuccessful cloning</td>
<td>Nasvi2EG007785</td>
<td>-0.951984</td>
<td>0.0007</td>
<td>0.0429312</td>
</tr>
<tr>
<td>LOC100116873</td>
<td>Ubiquitously and anteriorly expressed in pre-blastoderm stage unsuccessful cloning</td>
<td>Nasvi2EG011240</td>
<td>-0.71322</td>
<td>0.0001</td>
<td>0.00879605</td>
</tr>
<tr>
<td>total: 12 transcripts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Transcripts from the first experiment

<table>
<thead>
<tr>
<th>NCBI annotation</th>
<th>Name applied in manuscript</th>
<th>OGS 2.0 name</th>
<th>log2(fold_change)</th>
<th>p_value</th>
<th>q_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>coronin-2B-like (Coronin)</td>
<td>Nasvi2E001470</td>
<td>1.1236</td>
<td>5.00E-05</td>
<td>0.0048442</td>
</tr>
<tr>
<td>LOC100122439</td>
<td>kayak</td>
<td>Nasvi2E001004</td>
<td>1.2084</td>
<td>5.00E-05</td>
<td>0.0103145</td>
</tr>
<tr>
<td>LOC100116598</td>
<td>kayaking</td>
<td>Nasvi2E012316</td>
<td>1.1612</td>
<td>0.00015</td>
<td>0.0279964</td>
</tr>
<tr>
<td>LOC100122849</td>
<td>Endolucanase E-4-like (Endolucanase)</td>
<td>Nasvi2E005494</td>
<td>0.673426</td>
<td>0.0003</td>
<td>0.0511239</td>
</tr>
<tr>
<td>LOC103315681</td>
<td>weakly expressed in oosome, not in pole cells</td>
<td>Nasvi2E009650</td>
<td>0.941784</td>
<td>0.0004</td>
<td>0.0627122</td>
</tr>
<tr>
<td>LOC100120746</td>
<td>weakly expressed in oosome, not in pole cells</td>
<td>Nasvi2E003242</td>
<td>0.721212</td>
<td>0.0053</td>
<td>0.4616339</td>
</tr>
<tr>
<td>LOC100124030</td>
<td>ADP ribosylation</td>
<td>Nasvi2E007999</td>
<td>0.804803</td>
<td>0.0067</td>
<td>0.562907</td>
</tr>
<tr>
<td>total: 12 transcripts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.7.2 Command jobs.

The command jobs for analyses using the NCBI annotation 102 and the annotation OGS 2.0 with various normalization parameters.

15-11-23newnomergenew2.0gffcd+b (OGS 2.0 annotation)

#!/bin/bash

#MSUB -l nodes=16
#MSUB -l partition=ALL
#MSUB -l walltime=5:00:00
#MSUB -m be
#MSUB -o

export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11.Linux_x86_64:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0.Linux_x86_64:$PATH

echo $PATH

cd /mnt/lustre/jlynch42

tophat2 -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqP4th --library-type fr-firststrand ass1nvbowtie2
NvAp/Lynch12_CTTGTA_L008_R1_001.fastq NvAp/Lynch12_CTTGTA_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqA4firstth --library-type fr-firststrand ass1nvbowtie2
NvAp/Lynch11_GGCTAC_L008_R1_001.fastq NvAp/Lynch11_GGCTAC_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 --library-type fr-firststrand -o NvAp/6-11-15hhqA2firstth ass1nvbowtie2
NvAp/Lynch7_CAGATC_L008_R1_001.fastq NvAp/Lynch7_CAGATC_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqA3firstth --library-type fr-firststrand ass1nvbowtie2
NvAp/Lynch9_GATCAG_L008_R1_001.fastq NvAp/Lynch9_GATCAG_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqP2firstth --library-type fr-firststrand ass1nvbowtie2
NvAp/Lynch10_TAGCTT_L008_R1_001.fastq NvAp/Lynch10_TAGCTT_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqP3firstth --library-type fr-firststrand ass1nvbowtie2
NvAp/Lynch10_TAGCTT_L008_R1_001.fastq NvAp/Lynch10_TAGCTT_L008_R2_001.fastq

15-11-91strandedstringcd (OGS 2.0 annotation)
#!/bin/bash
l
#MSUB -l nodes=16
#MSUB -l partition=ALL
#MSUB -l walltime=3:00:00
#MSUB -m be
#MSUB -o

export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0:$PATH

echo$PATH

cd /mnt/lustre/jlynch42


#!/bin/bash
l
#MSUB -l nodes=16
#MSUB -l partition=ALL
#MSUB -l walltime=3:00:00
#MSUB -m be
#MSUB -o

export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0:$PATH

echo$PATH

cd /mnt/lustre/jlynch42

cuffmerge -p 16 -o DV/Nv/15-10-16DVoldstcm -g nv2clean.gff -s assnv1.fa NvAp/1ststrandstlist.txt
#!/bin/bash

#MSUB -l nodes=16
#MSUB -l partition=ALL
#MSUB -l walltime=3:00:00
#MSUB -m be
#MSUB -o

export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11.Linux_x86_64:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0.Linux_x86_64:$PATH

echo$PATH

cd /mnt/lustre/jlynch42


cuffdiff -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqA2firstth ass1nvbowtie2 NvAp/Lynch7_CAGATC_L008_R1_001.fastq NvAp/Lynch7_CAGATC_L008_R2_001.fastq

cuffdiff -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqA3firstth ass1nvbowtie2 NvAp/Lynch9_GATCAG_L008_R1_001.fastq NvAp/Lynch9_GATCAG_L008_R2_001.fastq

cuffdiff -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqP2firstth ass1nvbowtie2 NvAp/Lynch8_ACTTGA_L008_R1_001.fastq NvAp/Lynch8_ACTTGA_L008_R2_001.fastq

cuffdiff -p 16 -r 0 -l 10000 -o NvAp/6-11-15hhqP3firstth ass1nvbowtie2 NvAp/Lynch10_TAGCTT_L008_R1_001.fastq NvAp/Lynch10_TAGCTT_L008_R2_001.fastq

15-11-91strandedstringnnjcd (OGS 2.0)

#!/bin/bash

#MSUB -l nodes=16
#MSUB -l partition=ALL
#MSUB -l walltime=3:00:00
#MSUB -m be
#MSUB -o

export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11.Linux_x86_64:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0.Linux_x86_64:$PATH

echo$PATH

cd /mnt/lustre/jlynch42


./NvAp/15-11-9hhqP4firstnnjth-st
./NvAp/15-11-9hhqP3firstnnjnth-st
./NvAp/15-11-9hhqA4firstnnjnth-st
./NvAp/15-11-9hhqP2firstnnjnth-st
./NvAp/15-11-9hhqA3firstnnjnth-st
./NvAp/15-11-9hhqA2firstnnjnth-st

stringtie 8-11-15hhqA2firstnnjnth/accepted_hits.bam -p 4 -o 15-11-9hhqA2firstnnjnth-st
stringtie 8-11-15hhqA3firstnnjnth/accepted_hits.bam -p 8 -o 15-11-9hhqA3firstnnjnth-st
stringtie 8-11-15hhqP2firstnnjnth/accepted_hits.bam -p 8 -o 15-11-9hhqP2firstnnjnth-st
stringtie 8-11-15hhqA4firstnnjnth/accepted_hits.bam -p 8 -o 15-11-9hhqA4firstnnjnth-st
stringtie 8-11-15hhqP3firstnnjnth/accepted_hits.bam -p 8 -o 15-11-9hhqP3firstnnjnth-st
stringtie 8-11-15hhqP4firstnnjnth/accepted_hits.bam -p 8 -o 15-11-9hhqP4firstnnjnth-st


export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11.Linux_x86_64:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0.Linux_x86_64:$PATH

echo$PATH

cd /mnt/lustre/jlynch42

16-7-29HHQ2.1-1st-NOmerge-st-cd (NCBI 2.1 v102)

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqP31strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch10_TAGCTT_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqP41strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch11_GGCTAC_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqA31strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch9_GATCAG_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqA41strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch7_CAGATC_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqP21strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch8_ACTTTGA_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqP31strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch8_ACTTGA_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqA21strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch7_CAGATC_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqP21strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch8_ACTTGA_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqP31strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch8_ACTTGA_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/16-7-28hhqP41strandgenomic2.1th --library-type fr-firststrand -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch8_ACTTGA_L008_R1_001.fastq

18-2-26HHQ2.1-NOmerge-nnj-cd (NCBI 2.1 v102)

export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11.Linux_x86_64:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0.Linux_x86_64:$PATH
echo$PATH

tophat2 -p 16 -r 0 -I 10000 -o NvAp/18-2-21hhqA2genomic2.1th --no-novel-juncs -G Refseq_annotgff_Nvit_2.1_genomic.gff bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch7_CAGATC_L008_R1_001.fastq

tophat2 -p 16 -r 0 -I 10000 -o NvAp/18-2-21hhqP3genomic2.1th/accepted_hits.bam,NvAp/18-2-21hhqP4genomic2.1th/accepted_hits.bam

tophat2 -p 16 -r 0 -I 10000 -o NvAp/18-2-21hhqP2genomic2.1th/accepted_hits.bam
18-2-28APnotransmerged2.1-cd (NCBI 2.1 v102)

echo$PATH

cd /mnt/lustre/jlynch42


tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqA2notrans2.1th bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch7_CAGATC_L008_R1_001.fastq NvAp/Lynch7_CAGATC_L008_R2_001.fastq


tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqP2notrans2.1th bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch8_ACTTGA_L008_R1_001.fastq NvAp/Lynch8_ACTTGA_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqP3notrans2.1th bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch10_TAGCTT_L008_R1_001.fastq NvAp/Lynch10_TAGCTT_L008_R2_001.fastq

export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11.Linux_x86_64:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0.Linux_x86_64:$PATH

echo$PATH

cd /mnt/lustre/jlynch42


tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqA2notrans2.1th bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch7_CAGATC_L008_R1_001.fastq NvAp/Lynch7_CAGATC_L008_R2_001.fastq


tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqP2notrans2.1th bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch8_ACTTGA_L008_R1_001.fastq NvAp/Lynch8_ACTTGA_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqP3notrans2.1th bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch10_TAGCTT_L008_R1_001.fastq NvAp/Lynch10_TAGCTT_L008_R2_001.fastq
```bash
tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqP4genomic2.1notransth
t2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch12_CTTGTA_L008_R1_001.fastq
NvAp/Lynch12_CTTGTA_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqA4genomic2notranscript.1th
bt2Refseq_genomefasta_Nvit_2.1_genomic NvAp/Lynch11_GGCTAC_L008_R1_001.fastq
NvAp/Lynch11_GGCTAC_L008_R2_001.fastq

18-2-28HHQ2.1-transonly-cd

export PATH=$HOME/bowtie2-2.2.1:$HOME/tophat-2.0.11.Linux_x86_64:$HOME/samtools-0.1.19:$HOME/cufflinks-2.2.0.Linux_x86_64:$PATH

echo$PATH

cd /mnt/lustre/jlynch42
-u Refseq_annotgff_Nvit_2.1_genomic.gff NvAp/27-2-
21hhqA2transonly2.1th/accepted_hits.bam,NvAp/27-2-
21hhqA3transonly2.1th/accepted_hits.bam,NvAp/27-2-21hhqA4transonly2.1th/accepted_hits.bam
NvAp/27-2-21hhqP2transonly2.1th/accepted_hits.bam,NvAp/27-2-
21hhqP3transonly2.1th/accepted_hits.bam,NvAp/27-2-21hhqP4transonly2.1th/accepted_hits.bam

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqA2transonly2.1th -G
Refseq_annotgff_Nvit_2.1_genomic.gff -T bt2Refseq_genomefasta_Nvit_2.1_genomic
NvAp/Lynch7_CAGATC_L008_R1_001.fastq NvAp/Lynch7_CAGATC_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqA3transonly2.1th -G
Refseq_annotgff_Nvit_2.1_genomic.gff -T bt2Refseq_genomefasta_Nvit_2.1_genomic
NvAp/Lynch9_GATCAG_L008_R1_001.fastq NvAp/Lynch9_GATCAG_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqP3transonly2.1th -G
Refseq_annotgff_Nvit_2.1_genomic.gff -T bt2Refseq_genomefasta_Nvit_2.1_genomic
NvAp/Lynch10_TAGCTT_L008_R1_001.fastq NvAp/Lynch10_TAGCTT_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqP4transonly2.1th -G
Refseq_annotgff_Nvit_2.1_genomic.gff -T bt2Refseq_genomefasta_Nvit_2.1_genomic
NvAp/Lynch12_CTTGTA_L008_R1_001.fastq NvAp/Lynch12_CTTGTA_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqA4transonly2.1th -G
Refseq_annotgff_Nvit_2.1_genomic.gff -T bt2Refseq_genomefasta_Nvit_2.1_genomic
NvAp/Lynch11_GGCTAC_L008_R1_001.fastq NvAp/Lynch11_GGCTAC_L008_R2_001.fastq

tophat2 -p 16 -r 0 -l 10000 -o NvAp/27-2-21hhqP2transonly2.1th -G
Refseq_annotgff_Nvit_2.1_genomic.gff -T bt2Refseq_genomefasta_Nvit_2.1_genomic
NvAp/Lynch8_ACTTGA_L008_R1_001.fastq NvAp/Lynch8_ACTTGA_L008_R2_001.fastq

2.7.3 Primer list.
```
All primers used in this manuscript are listed. This list includes the primers used for cloning the genes and making templates for the probes and dsRNAs.

<table>
<thead>
<tr>
<th>NCBI annotation 102 name</th>
<th>OGS 2.0 name</th>
<th>Name applied in manuscript</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC100680515</td>
<td>Nasvi2EG022897</td>
<td></td>
<td>GGC CGC GGC ACT</td>
<td>CCC GGG GCC GTC</td>
</tr>
<tr>
<td>N/A</td>
<td>Nasvi2EG006979</td>
<td></td>
<td>GAT TGC CAA GCT</td>
<td>TGG ATG TGT GCT</td>
</tr>
<tr>
<td>LOC100121206</td>
<td>Nasvi2EG006980</td>
<td>mex3</td>
<td>GGC CGC GGG ACC</td>
<td>CCA G</td>
</tr>
<tr>
<td>LOC100114679</td>
<td>Nasvi2EG017534</td>
<td></td>
<td>GAG TGC CAA GCT</td>
<td>ACA C</td>
</tr>
<tr>
<td>LOC100119882</td>
<td>Nasvi2EG007110</td>
<td></td>
<td>GGC CGC GGG ACC</td>
<td>CCC GGG GCC AAA</td>
</tr>
<tr>
<td>LOC107982211</td>
<td>Nasvi2EG007111</td>
<td></td>
<td>AGA TGC ACC AGC</td>
<td>GGA G</td>
</tr>
<tr>
<td>LOC100122439</td>
<td>Nasvi2EG010044</td>
<td>coronin-2B-like (Coronin)</td>
<td>GGC CGC GGA TAT</td>
<td>CCC GGG GCC AAG</td>
</tr>
<tr>
<td>LOC100123633</td>
<td>Nasvi2EG003943</td>
<td></td>
<td>GGC CGC GGT TGG</td>
<td>GGC GAC TGG AAA</td>
</tr>
<tr>
<td>LOC100121288</td>
<td>Nasvi2EG003644</td>
<td></td>
<td>GGC CGC GGT CGT</td>
<td>TCA C</td>
</tr>
<tr>
<td>LOC100120121</td>
<td>Nasvi2EG011856</td>
<td>Exdl2</td>
<td>GGC CGC GGT CGT</td>
<td>CCC GGG GCA CTT</td>
</tr>
<tr>
<td>LOC100115727</td>
<td>Nasvi2EG028970</td>
<td></td>
<td>GGC CGC GGT TTT</td>
<td>TCC T</td>
</tr>
<tr>
<td>LOC100121774</td>
<td>Nasvi2EG013933</td>
<td></td>
<td>GGC CGC GGT TTT</td>
<td>CTC T</td>
</tr>
<tr>
<td>LOC100313502</td>
<td>Nasvi2EG002890</td>
<td></td>
<td>GGC CGC GGT TTT</td>
<td>TCC T</td>
</tr>
<tr>
<td>LOC103315431</td>
<td>Nasvi2EG022279</td>
<td></td>
<td>GGC CGC GGT TTT</td>
<td>AAA T</td>
</tr>
<tr>
<td>LOC100118628</td>
<td>Nasvi2EG013461</td>
<td></td>
<td>GGC CGC GGC GGT</td>
<td>AAA A</td>
</tr>
<tr>
<td>LOC100123856</td>
<td>Nasvi2EG004048</td>
<td></td>
<td>GGC CGC GGC GGT</td>
<td>GAC T</td>
</tr>
<tr>
<td>LOC100121321</td>
<td>Nasvi2EG013803</td>
<td></td>
<td>GGC CGC GGA GTA</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100118491</td>
<td>Nasvi2EG004082</td>
<td></td>
<td>GGC CGC GGT ATC</td>
<td>CCC G</td>
</tr>
<tr>
<td>LOC100301993</td>
<td>Nasvi2EG002296</td>
<td></td>
<td>GGC CGC GGT ATC</td>
<td>CCC G</td>
</tr>
<tr>
<td>LOC100114679</td>
<td>Nasvi2EG006980</td>
<td></td>
<td>GGC CGC GGC GGT</td>
<td>CCC G</td>
</tr>
<tr>
<td>LOC100119882</td>
<td>Nasvi2EG007110</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC107982211</td>
<td>Nasvi2EG007111</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100122439</td>
<td>Nasvi2EG010044</td>
<td>coronin-2B-like (Coronin)</td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100123633</td>
<td>Nasvi2EG003943</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100121288</td>
<td>Nasvi2EG003644</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100120121</td>
<td>Nasvi2EG011856</td>
<td>Exdl2</td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100115727</td>
<td>Nasvi2EG028970</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100121774</td>
<td>Nasvi2EG013933</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100313502</td>
<td>Nasvi2EG002890</td>
<td></td>
<td>GGC CGC GGT TTT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC103315431</td>
<td>Nasvi2EG022279</td>
<td></td>
<td>GGC CGC GGT TTT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100118628</td>
<td>Nasvi2EG013461</td>
<td></td>
<td>GGC CGC GGT TTT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100123856</td>
<td>Nasvi2EG004048</td>
<td></td>
<td>GGC CGC GGT TTT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100121321</td>
<td>Nasvi2EG013803</td>
<td></td>
<td>GGC CGC GGA GTA</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100118491</td>
<td>Nasvi2EG004082</td>
<td></td>
<td>GGC CGC GGT ATC</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100301993</td>
<td>Nasvi2EG002296</td>
<td></td>
<td>GGC CGC GGT ATC</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100114679</td>
<td>Nasvi2EG006980</td>
<td></td>
<td>GGC CGC GGC GGT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100119882</td>
<td>Nasvi2EG007110</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC107982211</td>
<td>Nasvi2EG007111</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100122439</td>
<td>Nasvi2EG010044</td>
<td>coronin-2B-like (Coronin)</td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100123633</td>
<td>Nasvi2EG003943</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100121288</td>
<td>Nasvi2EG003644</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100120121</td>
<td>Nasvi2EG011856</td>
<td>Exdl2</td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100115727</td>
<td>Nasvi2EG028970</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100121774</td>
<td>Nasvi2EG013933</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100313502</td>
<td>Nasvi2EG002890</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC103315431</td>
<td>Nasvi2EG022279</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100118628</td>
<td>Nasvi2EG013461</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100123856</td>
<td>Nasvi2EG004048</td>
<td></td>
<td>GGC CGC GGT GAT</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100121321</td>
<td>Nasvi2EG013803</td>
<td></td>
<td>GGC CGC GGA GTA</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100118491</td>
<td>Nasvi2EG004082</td>
<td></td>
<td>GGC CGC GGT ATC</td>
<td>GGC G</td>
</tr>
<tr>
<td>LOC100301993</td>
<td>Nasvi2EG002296</td>
<td></td>
<td>GGC CGC GGT ATC</td>
<td>GGC G</td>
</tr>
</tbody>
</table>

<p>| LOC100680515             | Nasvi2EG022897 |                             | GGC CGC GGC ACT | CCC GGG GCC GTC |
| N/A                      | Nasvi2EG006979 |                             | GAT TGC CAA GCT | TGG ATG TGT GCT |
| LOC100121206             | Nasvi2EG006980 | mex3                        | GGC CGC GGG ACC | CCA G |
| LOC100114679             | Nasvi2EG017534 |                             | GAG TGC CAA GCT | ACA C |
| LOC100119882             | Nasvi2EG007110 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC107982211             | Nasvi2EG007111 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100122439             | Nasvi2EG010044 | coronin-2B-like (Coronin)   | GGC CGC GGA TAT | CCC GGG GCC AAA |
| LOC100123633             | Nasvi2EG003943 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100121288             | Nasvi2EG003644 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100120121             | Nasvi2EG011856 | Exdl2                       | GGC CGC GGA TAT | CCC GGG GCC AAA |
| LOC100115727             | Nasvi2EG028970 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100121774             | Nasvi2EG013933 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100313502             | Nasvi2EG002890 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC103315431             | Nasvi2EG022279 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100118628             | Nasvi2EG013461 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100123856             | Nasvi2EG004048 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100121321             | Nasvi2EG013803 |                             | GGC CGC GGA GTA | CCC GGG GCC AAA |
| LOC100118491             | Nasvi2EG004082 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |
| LOC100301993             | Nasvi2EG002296 |                             | GGC CGC GGT GAT | CCC GGG GCC AAA |</p>
<table>
<thead>
<tr>
<th>LOC</th>
<th>Nasvi2EGID</th>
<th>Gene Name</th>
<th>Sequence 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC100678565</td>
<td>Nasvi2EG016467</td>
<td></td>
<td>GGC CGC GGC CTG ATG CCG TAC GCT TGG A</td>
</tr>
<tr>
<td>LOC100121187</td>
<td>Nasvi2EG009916</td>
<td>Nv-CG4552</td>
<td>GGC CGC GGG CTT CCT TGG TAG TGG ACG A</td>
</tr>
<tr>
<td>LOC100113954</td>
<td>Nasvi2EG027754</td>
<td></td>
<td>GGC CGC GGC TCC GTG TGG TCC TGA AAA A</td>
</tr>
<tr>
<td>LOC100115154</td>
<td>Nasvi2EG013231</td>
<td></td>
<td>GGC CGC GGT GAG GCA ACA GCA GAA AAG</td>
</tr>
<tr>
<td>LOC100122778</td>
<td>Nasvi2EG005059</td>
<td></td>
<td>GGC CGC GGG AAA CAG TGG CCT CAA TGA T</td>
</tr>
<tr>
<td>LOC100118857</td>
<td>Nasvi2EG015690</td>
<td>waterproof</td>
<td>GGC CGC GGT AGC AGG CCA AAG GAT CTG T</td>
</tr>
<tr>
<td>LOC100121645</td>
<td>Nasvi2EG008505</td>
<td></td>
<td>GGC CGC GCA TGG GAA GTC TAC GTC C</td>
</tr>
<tr>
<td>LOC100115713</td>
<td>Nasvi2EG012380</td>
<td>mad2</td>
<td>GGC CGC GGC TGC TGC AAG CTC AGA CAA G</td>
</tr>
<tr>
<td>LOC100118416</td>
<td>Nasvi2EG018583</td>
<td>milton</td>
<td>GGC CGC GGC CAA GCC TCG ACT CTA CAG G</td>
</tr>
<tr>
<td>LOC100117722</td>
<td>Nasvi2EG000058</td>
<td></td>
<td>GGC CGC GGC AGC AGG TTC CCA GAA CAG T</td>
</tr>
<tr>
<td>LOC100119116</td>
<td>Nasvi2EG006868</td>
<td>anillin</td>
<td>GGC CGC GGT GTT TGG CCG ACT TGT CAG C</td>
</tr>
<tr>
<td>LOC100119940</td>
<td>Nasvi2EG016344</td>
<td></td>
<td>GGC CGC GGA TGT GGC AAA CCA CTG C</td>
</tr>
<tr>
<td>LOC100115818</td>
<td>Nasvi2EG018250</td>
<td>spt5</td>
<td>GGC CGC GGG GGC GAC TCT ACA AAG CAA A</td>
</tr>
<tr>
<td>LOC100120350</td>
<td>Nasvi2EG010895</td>
<td></td>
<td>GGC CGC GGT CCC TGG AGG AGA ACA ACA C</td>
</tr>
<tr>
<td>LOC1000679945</td>
<td>Nasvi2EG016126</td>
<td>Nv-OoCLANK</td>
<td>GGC CGC GGC GAG TGG GAA CTC GGT CAT A</td>
</tr>
<tr>
<td>LOC100120333</td>
<td>Nasvi2EG015809</td>
<td>innexin1</td>
<td>GGC CGC GGG GAG GAC TTT ACA ACA ACA C</td>
</tr>
<tr>
<td>LOC100122980</td>
<td>Nasvi2EG005511</td>
<td>elbow</td>
<td>GGC CGC GGG GTC CGG AAT ATC TGA CAG C</td>
</tr>
<tr>
<td>LOC100120940</td>
<td>Nasvi2EG006450</td>
<td>mp1</td>
<td>GGC CGC GGA CAT CAA TCT GGG AGA CAG C A</td>
</tr>
<tr>
<td>LOC1000678249</td>
<td>Nasvi2EG016574</td>
<td>qin</td>
<td>GGC CGC GGC TGG AGC AAA TCC AAC CAC T</td>
</tr>
<tr>
<td>LOC100122778</td>
<td>Nasvi2EG005059</td>
<td></td>
<td>GGC CGC GGG AAA CAG TGG CCT CAA TGA T</td>
</tr>
<tr>
<td>LOC100118857</td>
<td>Nasvi2EG015690</td>
<td>waterproof</td>
<td>GGC CGC GGT AGC AGG CCA AAG GAT CTG T</td>
</tr>
<tr>
<td>LOC100121645</td>
<td>Nasvi2EG008505</td>
<td></td>
<td>GGC CGC GCA TGG GAA GTC TAC GTC C</td>
</tr>
<tr>
<td>LOC100115713</td>
<td>Nasvi2EG012380</td>
<td>mad2</td>
<td>GGC CGC GGC TGC TGC AAG CTC AGA CAA G</td>
</tr>
<tr>
<td>LOC100118416</td>
<td>Nasvi2EG018583</td>
<td>milton</td>
<td>GGC CGC GGC CAA GCC TCG ACT CTA CAG G</td>
</tr>
<tr>
<td>LOC100117722</td>
<td>Nasvi2EG000058</td>
<td></td>
<td>GGC CGC GGC AGC AGG TTC CCA GAA CAG T</td>
</tr>
<tr>
<td>LOC100119116</td>
<td>Nasvi2EG006868</td>
<td>anillin</td>
<td>GGC CGC GGT GTT TGG CCG ACT TGT CAG C</td>
</tr>
<tr>
<td>LOC100119940</td>
<td>Nasvi2EG016344</td>
<td></td>
<td>GGC CGC GGA TGT GGC AAA CCA CTG C</td>
</tr>
<tr>
<td>LOC100115818</td>
<td>Nasvi2EG018250</td>
<td>spt5</td>
<td>GGC CGC GGG GGC GAC TCT ACA AAG CAA A</td>
</tr>
<tr>
<td>LOC100120350</td>
<td>Nasvi2EG010895</td>
<td></td>
<td>GGC CGC GGT CCC TGG AGG AGA ACA ACA C</td>
</tr>
<tr>
<td>LOC1000679945</td>
<td>Nasvi2EG016126</td>
<td>Nv-OoCLANK</td>
<td>GGC CGC GGC GAG TGG GAA CTC GGT CAT A</td>
</tr>
<tr>
<td>LOC100120333</td>
<td>Nasvi2EG015809</td>
<td>innexin1</td>
<td>GGC CGC GGG GAG GAC TTT ACA ACA ACA C</td>
</tr>
<tr>
<td>LOC100122980</td>
<td>Nasvi2EG005511</td>
<td>elbow</td>
<td>GGC CGC GGG GTC CGG AAT ATC TGA CAG C</td>
</tr>
<tr>
<td>LOC100120940</td>
<td>Nasvi2EG006450</td>
<td>mp1</td>
<td>GGC CGC GGA CAT CAA TCT GGG AGA CAG C A</td>
</tr>
<tr>
<td>LOC1000678249</td>
<td>Nasvi2EG016574</td>
<td>qin</td>
<td>GGC CGC GGC TGG AGC AAA TCC AAC CAC T</td>
</tr>
<tr>
<td>LOC</td>
<td>Genbank Accession</td>
<td>Gene Name</td>
<td>DNA Sequence</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>LOC100678440</td>
<td>Nasvi2EG019633</td>
<td>rrm</td>
<td>GGC CGC GGA AGC ATC GTC AAT CGA ATC C GCC CGC GGT TTA ATC TTC CTC CGC GAA TTA</td>
</tr>
<tr>
<td>LOC100113617</td>
<td>Nasvi2EG009070</td>
<td>C16orf87</td>
<td>GCC CGC GGT CTA ATC TTC CTC CGC GAA TTA GCC CGC GGT CTC AGC TTT ACT G CCCC GGG GCA ATA GAT C</td>
</tr>
<tr>
<td>LOC100679293</td>
<td>Nasvi2EG009257</td>
<td>bark beetle</td>
<td>GCC CGC GGT CTA CTG CAC AAC GTC TCC A GCC CGC GGA GAT CGA AGG GAC ACA TTG C</td>
</tr>
<tr>
<td>LOC100120269</td>
<td>Nasvi2EG003216</td>
<td>coronin-2B</td>
<td>GCC CGC GGG CCG AGA ACA TCC TCT TCA G GCC CGC GGT CCC GGT CTC AGC TTT ACT G</td>
</tr>
<tr>
<td>LOC100122439</td>
<td>Nasvi2EG010044</td>
<td>kayak</td>
<td>GCC CGC GGG CCG GCC TAC AGA ACA TCC TCT TCA G GCC CGC GGG GTG ACC CTC GAC TAC TCA G</td>
</tr>
<tr>
<td>LOC100116598</td>
<td>Nasvi2EG012316</td>
<td>Endogluconase</td>
<td>GGC CGC GGG CCG GCC TAC ATC TTC TCT ATC C GCC CGC GGT CCC GTG AGG GTC AAG TTA</td>
</tr>
<tr>
<td>LOC100122849</td>
<td>Nasvi2EG005494</td>
<td>ADP ribosylation factor-like 4 (Nv-ARL4)</td>
<td>GCC CGC GGC AAA GCC AAG GGT GTG AAC T ACC CTC GAC TAC TCA C</td>
</tr>
<tr>
<td>LOC103315681</td>
<td>Nasvi2EG009650</td>
<td>N/A</td>
<td>GCC CGC GGT CCG GCC TAC AGA TAA C GCC CGC GGC AAG GCC AAG GGT GTG TGG T</td>
</tr>
<tr>
<td>LOC100680147</td>
<td>Nasvi2EG003874</td>
<td>ADP ribosylation factor-like 4 (Nv-ARL4)</td>
<td>GCC CGC GGC AAA GCC AAG GGT GTG AAC T ACC CTC GAC TAC TCA C</td>
</tr>
<tr>
<td>LOC100120746</td>
<td>Nasvi2EG003242</td>
<td>N/A</td>
<td>GCC CGC GGG CCA GCC TAC AGA TAA C GCC CGC GGC AAG GCC AAG GGT GTG TGG T</td>
</tr>
<tr>
<td>LOC100124030</td>
<td>Nasvi2EG000799</td>
<td>insulin-like growth factor 1 (igf)</td>
<td>GCC CGC GGC AAA GCC AAG GGT GTG AAC T ACC CTC GAC TAC TCA C</td>
</tr>
</tbody>
</table>

2.8 Literature Cited


CHAPTER 3

DISSECTING THE BEAN BEETLE *CALLOSObRUCHUS MACULATUS* GERM

PLASM ASSEMBLY MECHANISM IN THE CONTEXT OF TELOTROPHIC OOGENESIS
3.1 Abstract

Germ cells are specified by maternal determinants in *Callosobruchus maculatus* as in *Nasonia vitripennis* and *Drosophila melanogaster*. However, the germ plasm assembly occurs in the context of telotrophic oogenesis in *Callosobruchus* rather than the polytrophic oogenesis in *Nasonia* and *Drosophila*. In this project, we attempted to dissect the germ plasm assembly mechanism in *Callosobruchus* and examine the functions of the essential genes in germ plasm assembly and germ cell formation. We successfully cloned the genes and identified their expression patterns in the ovarioles and the early embryos. In addition, we found that *bruno* mRNA is localized in the anterior pole of the oocyte and early embryos, which is a novelty in *Callosobruchus*. We knocked down these genes by parental RNA interference and found their potential roles in germ plasm assembly. However, due to the limitations of the proper markers and the availability of necessary techniques, there are still gaps in our knowledge of essential gene functions to determine the mechanism of the germ plasm assembly in the telotrophic oogenesis of the bean beetle *Callosobruchus*.

3.2 Background

3.2.1 The maternal provision mode

The maternal provision mode, one of the two major modes for germline specification, relies on the germ plasm to specify the germline [2]. During early embryogenesis, the germ plasm associates with several nuclei and buds out of the
embryos to form the primordial germ cells. The PGCs continue to divide to produce more PGCs during blastoderm stage, and then they migrate to the gonad during gastrulation [2].

In the early 20th century, series of experiments were done on different insects that showed the cytoplasm in the posterior region of the embryo (the germ plasm) was necessary for germ cell formation [3-5]. Researchers also found that this specialized cytoplasm contains germline determinants that are synthesized during oogenesis. These maternal determinants are organized as the germ plasm, of which the morphology varies in different species. For example, the polar granules, the germ plasm in Drosophila, are in a form of static granules at the posterior pole of the early embryo [6-8], while in Nasonia it is an integrated large structure called oosome that can move around in the posterior region of the early embryo [9].

3.2.2 Germ plasm assembly in Drosophila and Nasonia

The best characterized model insect for germ plasm assembly and the maternal provision mode is Drosophila. The germ plasm assembly is a stepwise and oskar-directed process [10, 11]. In 1986 Lehmann and Nusslein-volhard showed that oskar mutant embryos lacked germ plasm and pole cells [12]. Later experiments conducted by Ephrussi and Lehmann in 1992 demonstrated that the oskar gene was sufficient to induce germ plasm assembly and germ cell formation [10]. They overexpressed the oskar mRNA and found that the amount of germ plasm increased, and therefore the pole cell number
increased, too. They also mislocalized the oskar mRNA at the posterior pole of the embryos and observed germ plasm at anterior pole, where the pole cells were formed later. These experiments showed that the oskar gene is both necessary and sufficient for the germ plasm assembly and germ cell formation [10-12]. In 2011, Lynch et al. found the Nasonia oskar ortholog in the Hymenopteran insect, the wasp Nasonia vitripennis, whose germ cells are also specified by maternal provision mode [9]. They showed that the oskar gene has the same function in the germline development as it does in Drosophila [9].

The oskar gene directs the stepwise process of germ plasm assembly during oogenesis in both Drosophila and Nasonia [9-11]. Oskar mRNA is synthesized in the nurse cells and bound by Bruno protein, which prevents the premature translation [13-15]. The localization of oskar mRNA in the oocyte is achieved by the polarized microtubule network [16-18] and by Staufen through mediating the interactions between the oskar mRNA 3'UTR and the motor proteins [19, 20]. When oskar mRNA reaches the posterior region of the oocyte, it is released from the Bruno and starts to translate Oskar protein [11, 14, 21]. After the initial localization of the oskar mRNA, Oskar protein is needed to bind to the mRNA and anchor it to the posterior pole of the oocyte. Absence of Oskar protein results in the detachment of the oskar mRNA in the posterior pole [20, 22, 23]. Therefore, the translational activation of the localized oskar mRNA is very important. However, so far there is not a unified theory that can explain the translational activation [24, 25]. Reveal et al. in 2010 showed that both clusters BREs (Bruno response elements in region AB and C) in the oskar mRNA 3’ UTR can repress the translation of unfocalized oskar mRNA and the C region BREs have an additional function in release the BRE-
dependent repression [24]. The other theory was proposed by Kim et al. in 2015, in which the mRNAs are translational activated because they are in the region of activation. This does not rely on the localization of the mRNA nor the functions of the BREs. The region-specific activation is in a graded pattern which includes the region where the oskar mRNA is localized [25].

Vasa and Tudor are the downstream proteins of Oskar [10, 11, 26-29], which help stabilize germ plasm and recruit other proteins and mRNAs, such as nanos and germ cell-less [20, 22, 30, 31]. The nanos mRNA is localized in the germ plasm and the Nanos protein forms a gradient in the posterior region, which represses the translation of the hunchback maternal mRNA in the posterior region of the embryo. Mutants in nanos caused the loss of abdomen segmentation [30-33].

The process of germ plasm assembly in Nasonia is similar to that in Drosophila. Lynch et al. identified the oskar gene in Nasonia and showed that oskar mRNA is localized both in the oocytes and early embryos [9]. They also used the Nasonia Vasa antiserum to show that Nasonia Vasa is also localized in the germ plasm. The knockdowns of oskar and vasa both resulted in no germ plasm assembly and no pole cell formation, and knockdown of tudor led to reduced size of the oosome and reduced number of germ cells, which might be due to the efficiency of the parental RNA interference [9]. Nanos mRNA is also localized in the Nasonia oosome, which also represses the translation of Nasonia hunchback maternal mRNA [34]. It is worth mentioning that a nanos ortholog has not been identified in the Callosobruchus genome, which makes the mechanism of posterior
patterning mysterious. In conclusion, the process of germ plasm assembly in *Nasonia* is similar to that in *Drosophila* [9].

3.2.3 Comparison of the ovary types

There are three different ovary types among insects (Fig. 3.1) [1, 35]. The panoistic ovary type is usually found in the basal insects, such as *Gryllus, thermobia, atrachya*, and *schistocerca*, etc. (Fig. 3.1., Fig. 1.2.) [35]. In this type of ovary, the germline cells become the oocytes, which are surrounded by the somatic follicular epithelia. There are no supporting cells during the maturation of the oocytes. All of the mRNAs and proteins have to be synthesized in the oocyte nuclei [1].

The meroistic ovary types are found in the higher insects and further divided into two types, the polytrophic ovary and the telotrophic ovary, both of which have the nurse cells that provide mRNAs, proteins and other cytoplasmic components for the oocytes. In the polytrophic meroistic ovary, each oocyte has its own population of nurse cells located at its anterior side and they engage in the direct connections via cytoplasmic junctions, which is found in *Drosophila* and *Nasonia*. In contrast, the telotrophic meroistic oocytes share a common population of nurse cells. Each oocyte has a nutritive chord protruding anterior to remotely connect with the nurse cells in the germarium region. *Tribolium* and *Callosobruchus* has this type of ovary. So far, the mechanisms of macromolecules transport from the nurse cell to the oocyte are still not clear, and are contributed by microtubules, microfilaments, muscular contraction, intercellular electrophoresis, and
Fig. 3.1. The schematic representation of ovarioles. (A) telotrophic meroistic ovary, (B) polytrophic meroistic ovaries (anterior-top; posterior-bottom). The germarium region, somatic follicle cells, support/nurse cells, and the nutritive chord are indicated within each ovariole chain. This figure is from [1].
osmotic forces, etc. [36]. Therefore, the mechanism of the transport can vary in different insects.

3.2.4 An evolutionary perspective on *Callosobruchus* germline

The representative of Hymenoptera, *Nasonia vitripennis*, and the representative of Diptera, *Drosophila melanogaster*, both have the *oskar* gene in their genomes. By examining the sequenced genomes of insects in the Holometabola and Hemimetabola, it appears that the *oskar* gene is limited to the Holometabola, and its presence corresponds with the use of maternal provision mode. This model changed with the discovery of an *oskar* gene in the genome of the cricket *Gryllus bimaculatus*, which does not play roles in germline development [37]. Thus, it appears that *oskar* gained a novel function when it was recruited as the germline determinant in an ancestor of holometabolous insects. However, not all insects use maternal provision mode in Holometabola, and accordingly these insects such as the silkworm *Bombyx*, the red flour beetle *Tribolium*, and the honeybee *Apis*, etc. lost the *oskar* gene in their genomes.

*Callosobruchus* and *Tribolium* are both in the order of Coleoptera, but they specify their germline differently. *Callosobruchus* uses maternal provision mode for germline specification, which is the same with *Drosophila* and *Nasonia*. As we know the germ plasm is assembled in the oocyte during oogenesis, and the nurse cells synthesize the necessary materials that are transported to the oocyte. However, *Callosobruchus* has a different oogenesis type from *Nasonia* and *Drosophila*. The ovary type in *Callosobruchus*
is telotrophic, whereas the ovary type in *Nasonia* and *Drosophila* is polytrophic. The major
difference between the two types of oogenesis is that there is a common set of nurse
cells that provision all oocytes in telotrophic ovaries, whereas each oocyte is fed by a set
of sister nurse cells that arise through incomplete cell divisions early in oogenesis. This
difference could make the transportation of the necessary materials synthesized in nurse
cells to oocyte as well as the microtubule network that mediates the transport very
different, hence the mechanism of germ plasm assembly in the context of telotrophic
oogenesis might be affected.

Since germ plasm assembly mechanism is well documented in *Drosophila* and in
*Nasonia*, dissecting the mechanism of germ plasm assembly in the context of telotrophic
ovaries in *Callosobruchus* is very important for understanding the maternal provision
mode and the evolution of the germline specification mechanisms.

### 3.3 Methods and Materials

#### 3.3.1 *Callosobruchus* rearing and sample collection

The generation time of the bean beetle *Callosobruchus* took 3-4 weeks in the 30°C
incubator (12 hours : 12 hours day night light cycle) with 30% humidity when they are fed
with blackeye peas, which need to be sterilized by going through -20°C for 24 hours and
then 60°C for overnight.
After being inseminated, the females lay eggs that are glued on the surface of the peas. The hatched larva burrow into the peas from the eggs and eat the peas. Before pupation, they burrow underneath the seed coat and later chew though the seed coat to emerge as adults. It takes 2-3 days for them to mature after they emerged. The adult beetles do not need food or water during their two weeks of adult lifetime. The new adults are separated from the old beans by going through the No. 5 sieve. They are transferred to the fresh beans for overnight egg lay. The beans with the eggs are collected and soaked on a shaking platform in the solution mixed from 50% embryo wash solution (0.7M NaCl and 0.05% Triton-X) and 50% bleach (Clorox) for 10 to 15 minutes. The eggs are collected by pouring the solution through the 120 μm mesh basket. The eggs are rinsed with water several times to remove the bleach, and then they are transferred into the fixative (2 mL 10% Formaldehyde, 2 mL 1x PBS and 5 mL heptane) for overnight fixation.

The females and males have distinct appearances. The females are usually larger than the males and have enlarged abdominal plate with dark strips on it, while the males have smaller plate with no stripes. To collect the ovariole samples, the newly matured females are separated from the males. Their abdomens are pulled apart in the PBS buffer. The ovaries contain mature eggs and ovarioles. The ovarioles are separated from the rest of ovaries. There is a membrane covering the ovarioles that must be removed for in situ hybridization and immunostaining. Once the ovarioles are collected, they should be fixed right away in the fixative (2 mL 10% Formaldehyde, 2 mL 1x PBS and 5 mL heptane) for around five hours. After fixation, the ovarioles are washed with PBS (supplement with
0.1% Tween 20) for several times and dehydrated with 100% Methanol. The dehydrated ovarioles are stored in the 100% Methanol at -20 °C.

To collect the cuticle samples, let the females lay eggs for overnight. The eggs on the beans are collected and put in the 30 °C for another 2 to 3 days incubation. The eggs are then collected using the same way described above, and they are transferred into the cuticle solution (90% lactic acid and 10% Ethanol) for at least two days incubation in 60 °C. After the incubation, the eggshells are carefully removed in the cuticle solution under the dissection microscope. The cuticles can be stored in the cuticle solution at room temperature.

3.3.2 Identification and cloning of the Callosobruchus genes

The transcriptome that was used to identify the Callosobruchus genes was published in 2016 [38]. The reciprocal blast was performed with Tribolium, Drosophila, and Nasonia. We identified oskar, bruno, vasa, and tudor, etc. The sequences of these genes are listed in the 3.6.

The protocols for making the probes and dsRNAs of these genes were adapted based on the protocols described in [39]. Since the Callosobruchus ovariole tissue is thicker than Nasonia ovariole, we added an extra permeabilized step in the protocol described in [40], which was using the Trixon-100 to make the ovariole more permeable.
A step of inactivating the RNAse was followed to preserve the RNAs for *in situ* hybridization. Detailed protocol is in Supplementary file 3.6.5.

### 3.3.3 Parental RNA interference on adult females

We use the parental RNA inference to knock down genes. The Hamilton Microliter #701 syringe and the Hamilton 90130 needle were sterilized in the 5% hydrogen peroxide solution for overnight and then were cleaned with nuclease-free water to remove the extra hydrogen peroxide. After the needle and the syringe were dry and assembled, dsRNA (1-1.5mg/μL) was loaded in the syringe. Since we used the female adults for injection and we targeted the ovary for gene knockdown, the injection point was at the vulnerable tissue between the hard-abdominal segments. We used the forceps to anchor the beetle and injected the dsRNA into the abdomen. The eggs were collected for cuticles since day 2 after injection and the ovarioles were collected at day 3 after injection.

### 3.3.4 The purification of the Oskar protein for generation of polyclonal antibody

In order to make the polyclonal antibody against Oskar, we purified the full length of the protein. We used the NEB pMAL™ Protein Fusion and Purification System (NEB #E8200S) to construct the fusion plasmid for protein purification. The coding sequence of *Callosobruchus oskar* gene was cloned and inserted into the pMAL vector downstream of the *malE* gene that encodes the maltose-binding protein (MBP). The MBP-Oskar fusion protein was expressed in *E. coli*. The *E. coli* cells were harvested and sonicated to release
the soluble fusion protein. The crude extract was running through the column with the amylose resin to capture the fusion protein. The resin was then washed several times with the column buffer to remove the unbound proteins. The fusion protein was eluted from the resin by maltose. The collected fusion protein was cleaved by Factor Xa to remove the MBP tag. After cleaving the soluble fusion protein, the Oskar protein was insoluble in the buffer while the MBP was still soluble. The insoluble protein was pelleted and denatured in 8M Urea and 3M imidazole solution. The denatured protein was sent to Pierce Biotechnology for immunization.

Two rabbits were injected with the denatured protein at day 1 and boosted at week 2, week 4 and week 8, respectively. The blood was harvested from the two animals for crude extract sera at day 56 and day 58. We got two samples of sera, PA7606 and PA 7607. Both sera were tested for immunostaining, only PA7607 serum gave specific and strong staining.

3.4 Results

3.4.1 Both oskar mRNA and protein are localized in the assembled germ plasm

The telotrophic meroistic oogenesis in *Callosobruchus* has the germline-derived nurse cells localized at the anterior tip of the ovarioles [41]. All of the oocytes in one ovariole share the same nurse cells and connect with them via nutritive cords [41, 42]. In contrast, the oocyte in polytrophic oogenesis has its own population of nurse cells at the anterior pole of the oocyte, and they connect with each other via cytoplasmic bridges [2,
Callosobruchus oskar mRNA and protein are detected at the same time and place in the oocyte, where the germ plasm has already assembled (Fig. 3.3). Therefore, their expression is not visible in the younger oocytes and in the nurse cells. This is different from the expression of the Nasonia ortholog during oogenesis, in which the oskar mRNA can be detected in the very early oocytes as soon as they are distinguished from the nurse cells and in the region of the nurse cells near to the oocyte [9]. In addition, Callosobruchus oskar mRNA is also localized in the germ plasm and pole cells during (Fig. 3.2 A, B), which is the same as Nasonia and Drosophila. The differences of the expression start time and the expression pattern in the nurse cells between Nasonia and Callosobruchus might result from the way the nurse cells and oocytes are arranged. Since germ plasm guides the germ cell formation in the early embryogenesis, the localization of oskar mRNA and protein in the germ plasm confirmed the germline specification strategy in Callosobruchus despite the differences in the oogenesis type and the evolutionary relations with Drosophila and Nasonia. As the germline marker, the oskar mRNA indicates the germ plasm shape and location in the early embryos. The germ plasm seems as integrated as the Nasonia oosome and is attached to the posterior cortex of the early embryos which is different from the mobile Nasonia oosome (Fig. 3.2). However, the observation might be otherwise since we were limited by the number of the early embryos that we could collect. It is worth to mention that collecting the Callosobruchus embryos is not as easy as collecting Nasonia embryos due to the thick eggshell, the sticky and cloudy yolk that make the embryos very easy to beak and hard to visualize for specific stages. Therefore, we need further examination to determine
Fig. 3.2 The expression patterns of the cloned genes in wild type *Callosobruchus* embryos. (A-B) *Callosobruchus oskar* mRNA is localized in the germ plasm (A, pre-blastoderm stage) and in the pole cells (B, blastoderm stage). (C-D) *Callosobruchus tudor* mRNA is localized in the germ plasm and in the migrating germ cells (D, gastrulation stage). (E-F) *Callosobruchus vasa* mRNA is localized in the germ plasm (E, pre-blastoderm stage) and in the pole cells (F, blastoderm stage). (G-H) *Callosobruchus bruno* mRNA is localized at the anterior pole of the embryos. (G, pre-blastoderm stage. H, blastoderm stage). The embryos are arranged with posterior side to right.
Fig. 3.3 The expression patterns of the cloned genes in wild type *Callosobruchus* telotrophic ovarioles. The ovarioles in this picture are arranged with anterior side to right. On the right side, the grey dots (white arrow heads) indicate the nuclei of the nurse cells, while the grey dots (white arrows) in circles on the left are the nuclei of the follicle cells surrounding the oocytes (yellow arrow heads). (A) the purple color shows the *oskar* mRNA localized in the germ plasm at the posterior pole of the oocyte. (B) The red color is the Oskar protein localized in the germ plasm at the posterior pole of the oocyte. The green color is the *bruno* mRNA localized at the anterior pole of the oocyte.
Fig. 3.4 The *oskar* knockdown results on ovarioles. The ovarioles in this picture are arranged with anterior side to right. (B, D, F, G, G') On the right, the grey dots indicate the nuclei of the nurse cells. (A, C, E, G) The grey dots in circles are the nuclei of the follicle cells surrounding the oocytes. The red color is the Oskar protein (Oskar polyclonal antibody). The blue color is the bruno mRNA. The green color is the immunostaining of the acetylated-alpha tubulin, which shows the microtubules. (A-B) The ovarioles collected from the females that were injected with nuclease-free water as the negative control. (C-G') The ovarioles collected from the females that were injected with the dsRNA specifically designed for *oskar* gene. (G') Closer look at the white box region on a different focal plane.
whether the germ plasm is static integrated structure that is attached to the cortex of the embryos.

3.4.2 Oskar is required for normal oogenesis and abdominal segmentation

We used the parental RNA interference to analyze the functions of Callosobruchus oskar during oogenesis and embryogenesis (See 3.3.3). The knockdown results showed that the Oskar protein was completely removed (Fig. 3.4 A, C, E), which suggested that the knockdown worked. Nasonia nanos is a good germline marker to indicate the germ plasm change after knockdowns. However, the gene nanos does not exist in Callosobruchus genome. Since Vasa and Tudor should also be localized in the germ plasm, we can use them as the germline markers. However, they are not available now. The Callosobruchus oskar knockdown affects the oogenesis but does not affect the bruno mRNA localization (Fig. 3.4 B, D, E, E’). The younger oocytes were not maturing (Fig. 3.4 B-E), which probably caused the accumulation of bruno mRNA in the very young oocyte (Fig. 3.4 E’), while oskar mRNA cannot be detected in the same stage’s oocyte (Fig. 3.4 A). Therefore, we cannot conclude whether Callosobruchus oskar plays the same role in germ plasm assembly as Nasonia and Drosophila oskar do. New germline marker need to be developed for further examination.

The knockdown effect on the cuticles exhibited the defects on the abdominal segments. The negative control that was injected with just water showed that there were ten normal abdominal segments (Fig. 3.5 A). However, the cuticles resulted from the
Fig. 3.5 The *oskar* knockdown results on cuticles. The cuticles are arranged with the head to left. The abdominal segments are labeled with numbers from 1 to 10. (A) This cuticle is from the egg laid by the female that was injected with nuclease-free water as a negative control. (B-D) These cuticles are from the eggs laid by the females that were injected with the dsRNA specifically designed for *oskar*. Yellow arrow heads indicate the segment fusion.
oskar knockdown showed certain extents of disruptions on the abdomen segmentation (Fig. 3.5 B-D). By comparing both sides of the cuticles, we found that some of them lost the segments on one side, which was segment fusion (Fig. 3.5 B, C, yellow arrow heads), and some of the severe ones just lost the whole segments (Fig. 3.5 D). This further verified that the germ plasm also has factors for embryonic patterning. In Drosophila and Nasonia, nanos is an important gene that is responsible for posterior patterning [30-33]. However, there is no nanos gene identified in Callosobruchus, which implies a different mechanism of posterior patterning.

3.4.3 Vasa and tudor mRNAs are not localized in the oocyte but in the early embryos

Vasa and Tudor proteins act downstream of oskar gene in germ plasm assembly. Vasa and tudor mRNAs in both Drosophila and Nasonia are not localized in the germ plasm, but their proteins are recruited by oskar mRNA and protein for germ plasm assembly [10, 11, 26-28, 43]. In both Nasonia and Drosophila, vasa and tudor are maternally and ubiquitously expressed in the early embryos. However, this is different in Callosobruchus who shares the same germline specification strategy as Nasonia and Drosophila. During oogenesis, the Callosobruchus vasa are expressed ubiquitously in the oocyte, especially in the younger ones, while tudor mRNA is not detected in the oocyte or in the nurse cells. In early embryos, both tudor and vasa mRNAs are localized in the germ plasm (Fig. 3.2 C-F). Vasa expression is extended in the pole cells (Fig. 3.2 F), and the expression of tudor is extended in the migrating germ cells during gastrulation (Fig. 3.2 D). We are not sure if vasa is expressed in the migrating germ cells as tudor is,
because of the lack of samples in this stage. It is also not clear whether the tudor mRNA is localized in the pole cells or not due to the same reason. Further characterization in these stages is needed.

3.4.4 Bruno is localized in the anterior pole of both oocyte and early embryo

The Bruno protein in Drosophila binds to the oskar mRNA during the transportation to the posterior pole of the oocyte to prevent oskar mRNA from being prematurely translated into protein [13-15, 44]. Based on the expression patterns from the BDGP in situ homepage, Drosophila bruno is maternally and ubiquitously expressed in the early embryos and it is not specifically localized in the polar granules. After the pole cells are formed, the expression extends to the whole embryo as well as the pole cells, then the rapidly maternal degradation depletes all the maternal mRNAs except for the ones in the pole cells. The expression in the pole cells persists in the migrating germ cells and later in the gonad. In Nasonia, the bruno is weakly and ubiquitously expressed across the stages. Callosobruchus bruno mRNA, however, is specifically localized in the anterior pole of the oocytes (Fig. 3.3 B) as well as in the early embryos (Fig. 3.2 G, H). This is a novelty in Callosobruchus, and its functions on germ plasm assembly and embryonic patterning are worthy of investigation.

3.4.5 Knockdown of bruno disrupted the localization of germ plasm and the oogenesis
Fig. 3.6 The *bruno* knockdown results on ovarioles. The ovarioles in this picture are arranged with anterior side to right. (B, D, F) The grey dots indicate the nuclei of the nurse cells. (A, C, E) The grey dots in circles are the nuclei of the follicle cells surrounding the oocytes. The red color is the Oskar protein (Oskar polyclonal antibody). The blue color is the *bruno* mRNA. The green color is the immunostaining of the acetylated-alpha tubulin, which shows the microtubules. (A-B) The ovarioles collected from the females that were injected with nuclease-free water as the negative control. (C-F) The ovarioles collected from the females that were injected with the dsRNA specifically designed for *bruno* gene.
We were curious to know the functions of *Callosobruchus bruno*, so we knocked down it by pRNAi. The ovariole samples were collected three days after the injection. The knockdown results showed that the integrity of the germ plasm indicated by Oskar protein was not disrupted, but the localization was flipped from posterior to anterior (Fig. 3.6 C, E). So far, it is not clear how this phenotype was generated. In *Drosophila*, Bruno binds to a specific region in the 3'UTR of the *oskar* mRNA to prevent premature translation, and Valois as well as other protein acts as the adaptors between a different 3'UTR region and the microtubule motors [21, 45]. Based on this, we proposed two explanations. The first explanation is the polarity of the oocyte was flipped. If this is the case, the mRNA would be transported to the oocyte from the nurse cells through the nutritive chords, and the mRNA was just localized at this “polarized” oocyte and translated into protein for germ plasm assembly. This would suggest that the *bruno* is crucial for the oocyte polarity and anterior-posterior patterning. The second explanation is the oocyte polarity was not disrupted. In this case, the *oskar* mRNA started to be translated once they reached to the joint between the nutritive chord and the anterior pole of the oocyte, where the germ plasm would be assembled. Further characterization on whether the anterior-posterior of the oocyte was changed is required to understand the role of *Callosobruchus Bruno*. In addition, the development of the younger oocytes (Fig. 3.6 C-F) were also arrested compared to the control (Fig. 3.6 A-B), because there were no other medium oocytes between the immature ones and the more developed ones.

3.5 Discussion
To elucidate the mechanisms of the germ plasm assembly and the germ cells formation, we used the pRNAi to analyze the gene functions by looking at the knockdown effects on the ovarioles, the embryos and the cuticles. However, due to lack of markers and the limitations on the techniques, we were unable to get all the necessary data to make a solid conclusion on how the germ plasm is assembled in Callosobruchus.

3.5.1 Markers are needed to study the functions of these genes

The lack of germline makers makes it impossible for us to examine the germ plasm after the knockdowns. Although the Oskar polyclonal antibody serves well as a germline marker, the Oskar protein was depleted after the gene was knocked down (Fig. 3.4 A, C, E). At the beginning, when we found that the Callosobruchus tudor and vasa were localized in the germ plasm of the early embryos, we predicted that they would also be localized in the germ plasm in the oocyte during oogenesis. However, this is not the case. The in situ hybridization results showed that they were not localized in the germ plasm of the oocytes. Instead the Callosobruchus vasa was ubiquitously distributed in the oocytes and the Callosobruchus tudor is not visible (data not shown). Therefore, we will need to find other germline markers. We tried the Vasa and Tudor antibodies from Nasonia and Drosophila. It turned out that they did not cross react with the Callosobruchus Vasa and Tudor proteins. The only option now is to make the antibodies for Callosobruchus Vasa and Tudor, since Vasa and Tudor proteins are the essential factors for germ plasm assembly in both Nasonia and Drosophila [10, 11, 46]. Therefore, Callosobruchus Vasa and Tudor should be localized in the germ plasm, which makes them the perfect marker
for the germ plasm. Other makers indicating the oocyte polarity and the anterior-posterior patterning are also needed to study the functions of *Callosobruchus bruno*.

### 3.5.2 Limitations of handling *Callosobruchus*

There are not many molecular experiments that have been done on *Callosobruchus*. Therefore, we need to develop necessary techniques and protocols before we can study the functions of the genes and the germline development. So far, we have developed the protocol for *in situ* hybridization and immunostaining on the ovarioles (Detailed protocol is in Additional file 1), which gives clear and specific staining. The extra steps we added were to use 1% Triton X-100 to further permeabilize the ovariole tissue which is too thick for just proteinase K to make it porous. As the tissue becomes more porous after the permeabilization, it was necessary to remove the RNAse by 0.1% Diethyl dicarbonate (DEPC). These steps guaranteed that the tissue was porous enough for the reagents to get in, and prevented the mRNAs inside of the tissue from being degraded.

Despite the progress of the molecular experiments made on the ovarioles, we still need to improve the embryo collection to get more and better embryos without eggshells. Since *Callosobruchus* embryos have very thick eggshells and sticky embryos, it is very hard to remove the eggshells by the cold methanol shock as we do in *Nasonia* and *Drosophila*. Hand dissection of the fixed embryos were tried to remove the eggshells, but this would usually lead to damages on the embryos. Therefore, we made little holes on
the eggshells to let the reagents come into the embryos for *in situ* hybridization. However, this did not give us best *in situ* hybridization results, and the DAPI staining that we used to indicate the developmental stages gave bad resolution though the eggshells. One of the feasible ways to solve this problem is to section the embryos and perform *in situ* hybridization and immunostaining on the coverslip. In this way, the embryonic tissue will be completely exposed to the reagents and the eggshell will not be in the way for visualization. However, the embryo yolk might still be auto-fluorescent, which will compromise the resolution. The same issue also occurs with the cuticle collection. The eggshell was hard to remove and interfered with the illumination from the dark-field microscopy, which made it hard to identify the details of the phenotypes on the cuticles. This problem was solved by dissecting off the eggshells. The other problem was that the tissue inside of the cuticle was not completely dissolved by the cuticle solution, and it also interfered with the lights from dark-field microscopy (Fig. 3.5). Other cuticle preparation methods should be examined to see if they can dissolve the tissue inside of the cuticle.

### 3.6 Supplementary materials

#### 3.6.1 *Callosobruchus oskar* cDNA sequence

5'TTCTTTTCAGTGTTATGTATTGTTTGATTACAACACTACCAACTCAGTGTATATGTTTTG
GCTACCAACTACCTAGGTACAACAGAAAAATGAGAAAATAATCATGTCTCTTTTAGG
GCCAGTAACTTAGTTTGGATCACGTTTCACTGTTTCAGTTAGTTCAGGTTTCC
AATTGCTCCAGAGACCTCTTCAGCTCTTCAGTCTCTCAGTCTTTCCATCATCACATTCA
CATTTGTAGGGAACATGAAAAACATGTTTTTCAATGTTTTTTTGTACTTTTTTCC
ACTCGCAGCAAAAAAATTGGCAGACTATATTTTCAGATCCAATTATACACGTCTAG
TAAGTATTTCTTGTATCAAAGGTGTAAACAAGCGAAGTTAGGGAGTGAATTCGTTTTG
ATATCAAAATGGTTTGGGGGACTATCAGAAATTACTGGAGGAGTTGGAGTGAATTCGTTTTG
TTCTTTTATCTATAATATGCCTGTACTGTGAAGACGGCCTTCTTCTAGACCATGTAG
ATGAAGAATTCAAATATGCCTGTACTGTGAAGACGGCCTTCTTCTAGACCATGTAG
ACTCAGCTTATATTGGCTTCTATGTGCCTTACATTACCAGAAATATTTATATTGTGCGTGATTGT
ACAATAGAAGATTATTTTGTATCCAGCAAATGTAATATATTTATTTGAGATGAAAGATGA
TTTTAAAGCAAAAAAGAAATGCTGCGAAAGTGCCACATGTAAGAATATTGACTTTCGAA
ACTACAACATTATCAGAACAAAGAGATATCAGAACAGACCATTCGGAAGGAT
CCTTTGTATAGCATCAATCAATAAACCTCTGACGTTATAGTTATGAAAGATTAGGA
CATTCTGCAAATTGGTGAATATGTGTCCTGACTTTTTAAACACATCGGCCTTGGAG
ATGACTTTTTTCCTAGATATAGCTGACAATACCTATTATGTACCTGTACCCG
GTATCAAGCAATGTGGCTAGGCTGCAAAACACTAGGATATTATGTACCTGATACGG
AGGGTGAAGAATACAGAAATGATAGCTCCAAGGGTGCTTCTGTAGGATTGATTTC
AGATTTGCTAATGGGCAATAACATAAACACATGATTATGTTTAAAGCAATTGAT
ACAGAATTGAAGAATAAGGCTGCTAGATTTACCTTGTCAACAATTCCACCGTCTCC
CAAGCTGCCTAATCCAGCAAATGTTGAGGCTCTGCATTTCATAATCAAGCCA
TTATGGACTATTCTTGGCTTTCTCTAGTTTAAATGCAATGTAATGAGCATAGATGAAT
ATTCCATCAGGAGCAATCCTACTTTTCAGAAGAGACTTTGACAGGCTCTCAAAGGTTG
CCAAAAACGTACATGATAGGTCTTCAGTGATTTATGGAAGGAAGATATTCTCAGC
ACTTTAAAGACTTGGCTTGAAGAAGAGGATTCAGGACATTGAGTGTCTCAGGAA
TACATACCTGTTCCTTCAGATATTTATATTATACTCAGTAATGTTTGTTGAGAATATT
GGTGAAGATATTTCACAGTCTTCACATGAATGTGAATTTTAGGGTGTTTAT
ATTTTGCAAGCTAATATAGCAAGATGTGATATTGTTGTTGTATTGTTTAAGTCAT
ATTGGATATTGATGCTAATATAAAATGATGAACAAACATATTTTGGTATGGGAATA
TTTGATAAATTCTTTTTATTGTTGTACAGAATATATTTTCGAAGAGAGTGAA
GTATACAGTTTTGATATGAAATATCTTTTATCAACAGATATGGAACCTAAAATATTTTT
GTAACATATAATTCTATGTTGGAAGTTAAAAGGCTGTAATGCTATGTAATAATTA
TCAGTGCTCAACAAGGTGACAGACATGAGTATATATGAAATGCCAAAATATTAAA
GACGAATACAGATTCTGGACTATGTCTATACACTGCTTTTTACTGCTCCACTATTACTC
CAAAAGTACATGAAACATAAAACTCAAAATTGTTGTAAGTTGGTCATTTGTGCTGGCAGTACA
ACTGAATTCAGTGCAATTACATAATTAAAGTTTTTAGTACTAATTACTCACACAGAGTT
AATTTTATAGAAAAATTAGGTATATTTTACTCTTTGTATCCGATATATTCTGTACCCCT
GATTTTTAAACTCAGGTAAGTTTTAAACAGAGAAAATGAAACACTAAGATTATTTTTAT
AGAAATGCTAATGCAACTAAAATTGTTCCAATAATACCAAAATATTGTTCCAGGCATT
TTTTAAACATCTGTAGATCTTAGCTGTAAGTTGATATTTATTAATTTATATATTAAGTA
TTTATCTAGTTTAAGATTATGAGAATGGAATTCAGGACAATATTGGATTCTGTGA
TATATTTTATTAGAAAAATCAACTCAACCCACTAATTATTTGTCTTTAAGTTTTATTCAA
ATAATCTTCTTACTTTTAATGCTGGGTCAATTGGTTTCAAATCTGGAAGAAATTCTG
AAGTTCACTTAGTTAAAATCAAGGTGACAAATTACTACTTACCACACTTTTTAT
TACATTAATGCTTTAATTTTCCCCAATATGGGTCTCAAACCTTACATGATGCTAT
ATTGATATATACCATTTAGGAGTTTGTCAAGACCAAGGGGACCCCAAAAGAAAAT
GAGGGGGGGGGGCGAAATAATCGTAAAGGTAATACAAAGTACC3'

3.6.2 *Callosobruchus tudor* cDNA sequence
5'CAATACAATAATAAATATTTAATACTGACGTATAACCTTTTTAACCAGATATAAAATG
AAATACACCTTTAAAAATCAACGAAAAACACATAAAATAGTAAACGCAGACTGGGA
TTTACCGTAATAATGCTTTTTGGGAAAAATCAGTTCAATTGCAATTTTGACCAGTCCA
TTGTAAAGGAGGTTCTCAAAAAGAAAAATTGTACTCAAAGTTCTAATTACTTCTTTGGTT
GAGATCTCGGAATTTTCTGTAAGGAATCTTCAGCTCCATTTCTCGATATGGAATTATCCA
GGCCCCACAAAGTCTCTTCACAAGGCCAGTTCAGCTCTCTCTGTTAGCAGTAATAACTGGT
GACATCTTTTGCGTCTAGGAAGTTCCAAACTGAGGTTCTGCGCCCTACTCAGCT
TTTGAGCAGTGAATATCGCCTGACCTTCTACGAGATATCTCCGAAACTTTGGTGGCC
ATCCTTTTGACCTTTGAAGTTCTCCTGCTTTTGGGAGGTTTCTCTTAAAGAGGACTTTTGGAGC
TGAGGCAGGCGTGTAGTTCTCTGAGCTCTCTGAGGCGGCGGTACCGTCC
CTGGCACACGAAACATTTCAACGTGAGACAAACATGTTTACGCCCAGCCACGGTGT
GGAGACTACCTGGCTCGTACCACATCATATCGTCGGGGAAAACGCGCAGCGCA
CAAGGTACCTGGCGAGTGATCTCTCAAGAAGTTTACTACCCGGTTGAGCAAGGT
GGCAGGATTTTTGGGCTCACAAATCCGGTAATTTCGTTCTGGGATTTGATTCA
CGGAAACCAGCAAAATTCTCTTGTGAGAATATCTTCTTGGACATTTTGGCCTCCAG
AAGGAGAGTTTTGGTGCTACAAATCCGGTAATTCGTTCTTGCCGATTGAATCA
CAGGAACCGCCAAATTCTCTCTTGTGAGAATATCTTCTTGGACATTTTGGCCTCCAG
TACAATTGACGTAAATTCTTTTTCTGCTTTGATCTAAATTCTGCGGAAGATTTT
CGACCAGGTCAACTCGCAAATCCGGTTCTTGCGCGCTTCATTCGACCATCT
GGCGGCGGCGGCTACAGTCTCTTGGTTCGCTATACAGCCAAATGACGAGTAGGT
TGAGGGCCAGGATTTCTCAGTGAGCTTTAATTGAGACATCCGGTATTGCAA
AAATCGACGAATCTGACGGAAGTAATGTCATCGTCAGCTCCAAAACTTTCCGGCT
GTACCACCTGGCATCTGCTGGAAGGAGACGCGCAGCCAGAAGCCTCGCGGCTTGGT
CAAAATCGACGTAAAAACACCTTATATTGATTGCCCACCATCACTTCTAGTACTTGAGCCC
TGTAaaaacatAttactttcagaaaacaacaataacTGGGACGGCAACAACTACAGA
GATTCAGAGTTTCGTCCTTTTATAGAAGACAGCTGAATATCTTCCATGAGGTTTTTGAAC
GGTGCCCTGACAATCTATCAGTCGACAGTAAAAATGAGCTGgGTTATGAAACCGGATTATCGCCTCCT
GATGGCTACATTGGAACACTGTCATCGACTACCTGAGCAGCTGATCCAGGAGCTCTTCATT
GACCGAACTTGGAAGTTTGTAGTGACGGTGCTTTTTGAAAACCTTTCTCAGAGCTCAGT
CCACCATAACCTCCTCCTCCTCCAGACATCATTTTTCTAAAGCGATTATCGCCTCCT
CCACCTCTGTCAATTACGGGAACTGTCGCTAGAGCCTTTTGGAAACTTTTTTCAAG
CTCGCTGGAAACTGACCATCACCACATTCTTTCTTATAGGGCCTGTTGAAACCGGATTTTGAACC
GTTCCATACGGTGCTCCTTGAACCGCCCTCCGTCGTTAGGAGACTCGTTCTTGAAGCC
GGTACCACCGTTCGATCGTTCTGAATCGCGCCGCCCCTCTCACAGTGCGACCCGA
GGCTTGCGGAAGGGTGCTCCCCTTCTTCGAAACCTCCCCCTTCGGTGCTGGTAAAGTTG
TCTGTTGCGAGTGTTGTCTTTCCCAGTTATCGGTTGGCGGAGATGTGGTATCGTT
CCAACCGTCCGCGTCTGTTGGGCACCTTCTCTCCTCCGCGGTGAATTTCGCGCGCTTCAG
CCGCCCGTGCTGGGCGTTGCCTGTCTCCTCCTTCGGTGGTATTGCCTAAACTC
CTTGTATATTTTCTCGCAAGAGAAGGGTTCTTTGAGTTTTTCTCAAGTTATCGTAGTT
TCGAGGTTCCTTGTCTGCGCCAGACTTCTCGTTTTTGTGTTTGCTGATTGTTGAA
ATTTTGTTGTCTCACGTCCAAGATTTTCGATCCCCCTCTAAATCTTTGCTCTCTATT
GTTCGGAGTGAGGCGATTCTCTTAGGTCGATTTGTGCTGTTCTGCTGGTTACGCA
GAACCCGGCTAACCTTTGGAGACTTTTGCGCCGCCCAAGTATCCAAATAACAGAGATGC
CACGTTGTagTTGGATTCGTCGTTAAGACTCGACCACGCGCTTTTCTTTGAGACATT
CGACCCACTTTGATTGAAAGGTACCTTTCTAAAATGATTTTCTTTCTAATAAGGCTGTC
TTTCAGCGTCTTCAACCATGTTCTGGTAACCATTCAACAGCATTCAATGGCTTTG
GGTCTCAAAACCGGTGCACTTGTTTACCCGTCACTGGATTATTTTAAGTTGAACAACTGGACAC
CACTGCTTCATTCCCATATAATATATTCGACACCTCTTGCTGCTGCTACTATAGAGCGACAGCA
TTTCAACGATTTTCAGCTCGGTACCATTGGCCCTGTCGCTTCTCAAATAGAGACACAGCA
GGCAAGCCCACCTTGATAGCACCACCTGAGTATGTCATTCACTTGACACGATTG
CTGAAGTTCTTGGTACAAGACTCGCTAGTTCATTTGCTTCTTTTGCTTACTAACTGAAAGTA
AAATCTGTTACAGCTGAAACGTAACCTATCTTGACTGGCTGACTGAAAGCCCTGTTT
CAGATGGCAGGACTTTTCATATAGTAACAGGCCTGGTTATCTACAAAATTTTCTGAAAGGCA
CACTGCATTTCTAAGTGAACAGTAGATTTTTCTACACCCTAAAGTGAAAATCTGATA
GCCATGACTTTTGGAAGGACATATTGAAAGGTATCTGATATAGAAAGTCCAAGG
TACACTTTCTGTGATGTTCCAAAATCTACATAGAATACCTTAAACTGATTATCGACTTC
ATTTGTCAACCACAGCCCTGCAATATGTTACCATCTCTCTGGCACCTAGCAGACAGA
TAGTGCCAGGAATTTGGAACATCTTCCAGAGGCCTCAATGGATTTGAGTTGATGCT
TTCATTAGCATGACCAGGACCTCTCCTTCACCTCTGTCAGTTGCACGGAAAAATGGCA
GGGTCCGTCGTGACATAGGATACATAAAAACTGGGATACAGGCAACCCAGGTTCCAAA
GTGAGCGCTTACTATAGTTAAGGCTCCCTCCACTCAGATTGTTGCTGCTGACAC
TGGTTGTGTTGGTGGCTTGTGGTTTGCTGATGTCATGCTCAGCATCAGCGCT
GGTGAAGCTTGCAATGAAATTTGATGTGCTGAGTTGATGGGAC
AGATCGGTTGTTTCAGATACAGCCTGGCCAGGAATCTGAGGTGACGCTGCTCGTAA
CAGGTTGCCAGCAGCAGCTCCCCGGAACTTCCACGTCTGCTTGGAGATGGGATAGGATAG
GGCCTGTTTCATGGTCACCACCACCACACTGAGCTGACCTCTGGCACAAGATG
CGCTAATGGTGGTACAGATATCAGTGACGCGGGAAAGGCCTGGAGACTTCTGATGTTCTCACACGGCACCACGTCTTTGTTGCCATAGTCTAAAAAATGACCTCCAGAAAACCCTCGTGCAAAAAGCTTGTGTTTATGATTCTTGCTCGGTAATATTTATTTGTCCTTTATATTGGCGCAAACCAGTGTT

CCAACATGCATGTTTTCCGGTGTCATCTTTCCTGC

TCCCAATTCGAACTGCGGACTGCTGACACCAGAAAACGTGGTCTACGTACGTAGCAC

TGCTCTTCTTTGTTGGCCCATATCTTCAAAAAAGGACCTTCTCCTCTTTTCCAGGTGC

GTGATGTACAGATTAAAGATATTATGCGGAAGATCCATCTAACTCAGTAG

TATCACTTATTTTGTGTTTTATAGCCTTTTTTCTTTGATTATTTGTAGTTCAAACCTGTAAACCGCGTGCTGCTCAGAACATCCTTCTCGAGAACAAGAAGCATTAG3'

3.6.3 *Callosobruchus vasa* cDNA sequence

5' TTTTTTTTTTTATATAAAATACCTTTACTGAGTGTTTCTACACCATAAACAGATTACT

AATTTAGTATCACAGCGTGTTATATGGCCTTCTGTAAGATCTCAAGTTAAACTTTTAGAA

TCAAAAACCTGGTTCTTAACATGTCAAAAACCTTAAACAAAAACTTCAAAACCAACCGCTT

ATTAAACACACAGGTAAACAAAAATAGAAACACAAAAACATACTTTACATACATAAAAC

CATGAACACCTGCATTGTCAACTGGATTGTCCTTCTAATTTATGGTAAGGCTTTTTTG

GATAACATTCAAGACCGGATTCAAAATTGATTCAATTTGGTCACTTTATATCTTGAA

TGTAATATATAACTTAAATAGTTAGTGTGTGTCAGGGTTTTTCAAAAAAAACAAAAATTGGT

TAAACCGATTATACACAGATAGTTTTAAGCCTCATAGACAGTGACCGGTTTCAAAATA

CTAACTAAAATTGATAAAGAAGCATAGTTTGGTTTTCTGTAATCTCATGAAACAGGATACA

GTTTATGATGAAACTAAAACAAAACACCTTTAAACAATAGAAAAAACAAAACAGTATTCTGTTAC

TTGTTTGTCAATAACATCTGCAGTGGTATGTTGGAAGAACCAGGTTTCTTACCTGAGAT
ATCAAAATCCTCTTTAAAAATTAATTACATTTTTTGTAATAACTTTTTCGAGCTCAAAAAGTG
AATGGTTCAAAAAATGTTTTATAATTATTAAGATCTTTTATATTAGTAAATACTAAAATAT
ATCTTAGATGTGTCTTTTTGATGGGAATATACCAACAAAAAGTATGTAAAAACGATCT
AGTAATAAGGCATTCTTTTGTTAATTGGAATCATCGTTTGATTAGTACTCTGTGGGC
TTATTTTGATTTTGTGCAAGTGCAACGAAACATGAATCTTTGGGTAAATCCGCTCCTCAAC
GTGAGCAACTTAATCTTATGAATTACACTCTTTCTCTGCTCTCTGCTCTGTGTTA
CAATGGGTGTTACTGCTTTCCCCAAAGTGCCGCCTCGTATATCGCGACCTCCGAAC
CTATCGGTGCTCCGCCTCCAGAGAATCTCCTCTCTCTCTCTACCAACCAATCGGGGTGCTC
TTTGACCAGCCCTGTTTCAAGATTTTTCGCAAGGCCTCCGCGCAAGGTTGCTCTGTGTT
CGGGGTGCAAAAAGCTTTTGCTTTTTCCTCCGGTTACCAACTCTCCCGTCCTTCCCG
ATACGGGTGGACGTACTCGTCTATGCTCTTTTGTTAATATCGTATTGATGACGTGGT
GATGCCCTCTCTATGTCCAGCCCTCTGGCTGCCACACCGGTGGCCACCAGGTGTTG
CATATGCCCTTTCTTAAAGTCCCACAGGGCCTGCTCCTCTTTGTTAATCGGTCT
GCCGTGTATGCTCGTCTTTGATGCTTTCTCGCTCAACAGGGGTTGCGCCAAAAGT
CTGCCGTTCCGTCTTTGTCTCGAGCAGACACCACTGTTTCTCACCAGCACCTTCTTG
AGCATTTCAATCAACTTCTCCTCTCTCTTGAACCTCGTGACTTCATGGACACCTGC
TCAACGTCGGTACAGGCGCTACCAACTATACACTTACAGCTACGAAAAACGTAGTTATT
TAAAAATTTTCCGCTAAATGCTGAGAATTCTCCAGGGAATGTAAGCTGAAAACATCAG
CGTTTGACGGTCAACCAGTGCAACCATAGTCCCATCCTACACCTCTTCTCTACTCGT
TTGGTAAGAAGCCCATGTCTCCTCAACAGCCATCGTGCTTTGCTGCAAGGAAAGAACTTG
ACAGATGCAGATGCAGAAGCTCTTCTCAGAAGTCATCCAGAGTCGTCAGGAGG
AGCCACAAGGATGTCGACGGCCAACACTGCACTATTTTCTTGATAACCTACAGACG
TGCCTCTCGTATGCAACACAGTCTTGATGTAACCTGCCATATGCAAACCTTTTTCGCT
GTTCGAAGGATCTGTATGGCGAGCTCCCTTGGTGCGATATGATTAATTACCAATGGC
TGACAGCAGAGATCTCTTCTACTACGACAGAGCCCGGTCTGTCTCCAATAATTGGTTGAT
CATTTGTTAGCAAAATTGCGACCGAGCTTTTTTCAGGACACGGTTTGGAGCACAACCCCATCA
AATCAGCAGCAGAAATTGATCCGCAGTACTTTTTGAATTGGATGTCAGATTGTTTTC
GTATATCCTGATTTTTTCACGTTTTTCTAACAGGTGAGCCCTAAGGCGAGGTAGGGTATC
CAGTTTTGACGAAATTGATCCGCAGTACTTTTTGAATTGGATGTCAGATTGTTTTC
TTGGTCCTTTTCTGAGGACACATATATCTCTCTGGTGACGGGCCTTTGCCGCTCTTTTG
CCGCCGCGTGCCGCGATCTCCGCTCTGTCGCGACCCGAAATTGCGCCCTCCTCTGCCCC
CACCAGCAGCGGCTCCGACCTGCTTGCCTCCTGAGCAGGACGGTTTGGCTCCGAGGCTCC
TACCGCTACTCAGCTGCCTCCGCCGCCCGAGTGCTTGGGTCTCCGCGGAGCTCC
ACCGCCCCAGCCGCTATCGCCCCCGCCGACGGGCAGGGCTTGGGCCTCTTTGCTCA
CTCTTAGAACCAGCCCCACCTCCGCCCGCGAAGCTGTCCGCTCGGCTCCGCGGCTGTC
AACGCGCTCCGCGCAGCTCCTCTCCGCGCTTCCGCGCCACCTCGACGCGCCG
TCCGCTGCAGGCAGCTCCCCCCCCCGGCTCGGCTGTCCGCTGCGCCGTCCG
AACCAGGCGCTCGCCCCACCTCGCTGCTGCTCCGCGCCGCTTCCGAGACCGCC
TCCGCTGCGCAGCAGCTCCCCGAGCGCGCTCGGCTGCGGCTCTCTCAGGTCTTTG
AACCAGCGACGCCGAGAAATTCTTACCACCTGCGATGCAGCGGCTACATTGCG
CCGAAGGTGTGACTGTGCAACCTGTATCATCCATTGATCATCCATGGCTATTATT
GTTGTAATTTTTCTCTTGTTAACCTGAAATTTTCTAGTAGATAAAAAATTTTACCCG
GTGAAGAGATGTCTCTGCTGCGTGGTAAGAGTAGGAAAGAAGGAAAATAATGG
ACAACGAACAAG3'
3.6.4 *Callosobruchus bruno* cDNA sequence

5’AACAACAAACAACAAACAACACAGCGTAGTTAGTTTGCAGCAGTCTCCGACTCCGA
TGAAAAATTGACCAGGAGCAGCCAGATAGTGATACAATAAAAAATGGTTTGTGTGCA
GTGCCCCGATCGATGGAGCCAAATGACCTTGAGGAGAGATTGAGGAAATTGAGCA
GGGTGCACTCGATAAACGTATTGAGGGATAAAGAACACAGGGCGAGTAAAGGTTG
TTGTGTTTGTACGTTTTTTACGCAAAAGCTGCCCTCCAAGCTCAAGATGCCTACA
CAACGTGAAGACGTAAACGGGATGCACCACATCTATTCAAGATGAAACCCGCTGCA
GGGAAATAGGAAGTGAGCAACAGCTGTTCGCAGGCCACGGCCGCACTCGAGGAGAAT
AGTACTGCCGCAGCGAGCAGCCGCACTCGGCAGCTGCCTTTGTACGTTTTCGC
CAGCAAAACAGAGCGCCCTAGGGGCGATCAAATCCCTGCAACCAGAGCCAGACTATG
GAGGGCTGTCTGCCCTGCTGCAATATCCGCACTGACAGGAGGAGAGGAGGAGT
GAATCTCAGACAGCAACAGATGCAGGCAAAAGCTGGTGTCGGCCGCTGGGGCG
CCCGCGCTCTCTCTCCTCCCGCCCGCACACCCAGCGTTCTCGCCGGTGCTGCGCTGCG
CCGCCGAGCTGCCGGCGGCCGCCGCAAGCAGCCTGCAAGTTGCAAGGGGTCG
GGCGTTTCCGGCGTCTCAGCAACAGCTGCTCCAGCCGGGCACCACCAGCAGGC
CTGCTCGCGCATTGGGGGCTAAGTACATCCCGTGATGAGATGGTCAGCCGAC
TGATGACGTGGCCGGCACCAGTGCCACCAGCCGGACCCGGACCCGGACCCGGACGC
CGCTGGTCTCCCTGGAGGCGCGCAACGGGACCAGGGGCGGGCGGGACGGACGC
CGTGGCTCGAGCAGGTGGGTTGCAGCGCAGCTGGGGTCGGCTGCGGCGCGGC
CGCGGGGGCG3’
3.6.5 FISH/IF protocol specific for *Callosobruchus maculatus* ovarioles

**Materials**

Before Day 1 the ovariole samples were collected and stored in 100% Methanol at -20 °C based on the description in 3.3.1.

**Day 1**

1. Rehydrate around 10 ovarioles through a methanol/PBT series (75%, 50%, 25%). Each time of rehydration needs five minutes on a shaking platform.
2. Wash in PBT three time. Each time is five minutes long on a shaking platform.
3. Remove the PBT from last step as much as you can. Add proteinase K in PBT for a final concentration 4 ug/mL and digest the ovariole samples at room temperature for 5 minutes. After the digestion is done, rinse the ovariole samples with PBT twice to remove the proteinase K.
4. Remove all the PBT, and then add 1% triton X-100 in PBT to the ovarioles. Put the tubes on ice for 30 minutes.
5. Wash the samples with PBT for two times on a shaking platform. Each time takes five minutes.
6. Incubated the samples with 0.1% Diethyl dicarbonate (DEPC) in PBT for two times. Each time takes 15 minutes. (Note: DEPC is hazardous, make sure to use proper protections.)
7. Wash the samples with PBT for two times on a shaking platform. Each time is 5 minutes long.

8. Fix the samples with 5% formaldehyde in PBT for 30 minutes.

9. Wash the samples with PBT for four times. Each time takes 5 minutes.

10. Incubate the samples with 50% hybridization solution and 50% PBT for 5 min at room temperature on a shaking platform.

11. Remove the solution from last step as much as you can. Add hybridization solution to the samples and incubate at 60°C for two hours.

12. Remove the hybridization solution and add the specific dig-labeled probe to each reaction with 1:50 ratio in hybridization solution. Incubate the samples at 60°C for overnight.

**Day 2**

1. Wash the samples in hybridization wash solution for three times at 60°C on a shaking platform. Each time is 25 minutes long.

2. Wash the samples in 50% MABT and 50% hybridization wash solution for 5 min at room temperature on a shaking platform.

3. Wash the samples with MABT solution for four times. Each time takes 10 minutes.

4. Block the samples in the blocking solution (10% western blocking solution in MABT) for at least one hour at room temperature on a shaking platform.

5. Incubate the samples with 1:50 anti-dig-POD antibody (1:1 in glycerol) for TSA detection in Day 3. Add other primary antibodies for immunostaining in Day 3 if needed.
Put the reactions at 4 °C for overnight. (Primary antibodies including anti-dig-POD are diluted in the blocking solution, use 500 uL for each reaction.)

**Day 3**

1. Wash the samples in MABT for four times. Each time is 15 minutes long.
2. Rinse the samples twice with PBS. For TSA detection, we use the Alexa Fluor™ 647 Tyramide SuperBoost™ Kit, goat anti-rabbit IgG (Thermo Fisher Scientific)
3. Dilute 30% H₂O₂ in 1:200 ratio in the amplification buffer. For one reaction, add 1uL of above dilution to 100uL amplification buffer for a final dilution of 0.0015% H₂O₂. For one reaction, add 1uL tyramide stock to above solution and immediately mix well and add to the samples.
4. Incubate the samples in dark for one hour at room temperature.
5. Wash the samples with PBT for three times. Each time is 5 minutes long.
6. Block the samples in the blocking solution in PBT for 30 minutes at room temperature.
7. Dilute the secondary antibodies in blocking solution and add it to the samples. Each reaction is 500 uL. Incubate the samples at 4 °C for overnight.

**Day 4**

1. Wash the samples in PBT for three times. Each time takes five minutes to remove the antibodies.
2. Add vestershiled mounting reagent with DAPI and put the samples at 4°C for overnight. The samples are ready for visualization afterward.

3.7 Literature Cited


CHAPTER 4

KNOCKING IN *CALLOSOPRUCHUS* OSKAR GENE IN THE GENOME OF

*TRIBOLIUM* USING CRISPR/CAS9 AND GAL4/UAS SYSTEMS
4.1 Background

The zygotic induction mode is the ancestral mode for germline specification, while maternal provision mode is a derived mode from it [2-4]. The origin of the maternal provision mode is from the common ancestor of Holometabola [2-4]. However, some of the Holometabolous insects have lost the ability of using maternal provision mode over the course of evolution such as Tribolium. They went back to use zygotic induction mode, instead of going for a third mechanism for germline specification [4]. This suggests that there might be a core regulatory network of the germline specification between the two modes.

Research on Drosophila germline specification had showed that mislocalization of the oskar mRNA at the anterior pole led to germ cell formation at anterior, and only three genes are essential for ectopic germ cell formation [7]. These three genes are oskar, vasa and tudor. In addition, the Drosophila needs necessary genes for germ cell formation, which are oskar, vasa, tudor, cappuccino, spire, staufen, valois and mago nashi. The presence of their Nasonia orthologs verify this conclusion (Checked in NCBI for orthologs, [3]). Interestingly, all these necessary genes for germ cell formation except for oskar gene are also found in Tribolium, an insect known for specifying its germline by inductive signals during blastoderm stage. So far, we have only examined the expression patterns of tudor, vasa and mago nashi in later embryogenesis stages in Tribolium. Vasa and tudor are ubiquitously and maternally expressed in the early Tribolium embryos, while they are zygotically expressed in the PGCs when they are being specified.
The prerequisite for maternal provision mode is to have fully assembled germ plasm during oogenesis. Although these germline-required genes are present in the *Tribolium* genome, and are expressed as mRNAs maternally, it seems that they are not assembled as the germ plasm. As we know the *oskar* gene sits on the top of the hierarchy for germ plasm assembly during oogenesis, it makes sense that these factors are not assembled during oogenesis because the *Tribolium* genome lacks the *oskar* gene. Given the conserved features between the two modes and the role of gene *oskar* for assembling the maternal determinants, it is logical to wonder whether the artificial provision of an *oskar* gene to *Tribolium* would show any sign of initiating the germ plasm assembly, which would represent the reversion to the use of the maternal provision mode for germline specification in *Tribolium*. Therefore, we think it is interesting to ask what would happen to the germline development when *oskar* gene is supplemented for maternal expression.

The *Callosobruchus oskar* gene the perfect candidate for this study. *Tribolium* and *Callosobruchus* are the closely related insects in the order of Coleoptera, and they share the same oogenesis type, the telotrophic oogenesis. The only thing different is that they use different strategies to establish their germline. In *Callosobruchus*, germ cells are specified by maternal inherited determinants, while in *Tribolium* germ cells are specified by inductive signals dependent on the zygotic genome. Therefore, knocking in the *Callosobruchus oskar* gene into the *Tribolium* genome for maternal expression is the best solution to test whether *Tribolium* has a latent capacity to assemble germ plasm.
Fig. 4.1 The genome structure of *Tribolium eagle*. The *Tribolium eagle* gene has three exons and two introns.
Fig. 4.2 Identification of the CRISPR target sites. (A) The sequence of the first exon of the *Tribolium eagle*, where 21 CRISPR target sites were identified. The red box indicates the two target sites that were chosen for sgRNAs design and production. (B) A close look at the two CRISPR target sites and their PAMs.
Fig. 4.3 The expression pattern of Tc-eagle in early embryos. (A) and (C) show that Tc-eagle is maternally expressed and localized at the anterior pole in pre-blastoderm stage. (B) and (D) are the DAPI staining to show the developmental stages. This figure is from [1].
In this project, I first attempted to use CRISPR/Cas 9 system (clustered regularly interspaced palindromic repeats/CRISPR associated 9) to directly knock in the coding sequence (CDS) of the *Callosobruchus oskar* gene into the *Tribolium* genome. However, this might be hard to maintain a stable transgenic line. Therefore, I combined the CRISPR technique with Gal4/UAS system to maintain two separate transgenic lines that essentially will not affect the normal growth. Since *oskar* gene has to be expressed and localized in the oocyte during oogenesis, we planned to insert the Gal4 gene into the first exon of *Tribolium eagle* gene in order to use the upstream regulatory elements and 3' UTR sequence (Fig. 4.1, Fig. 4.2, Fig. 4.3). Because *Tribolium eagle* is expressed maternally and localized in the anterior pole of the oocyte (Fig. 4.3), Gal4 protein would also be maternally expressed if this experiment works. We named this line the CRISPR_Gal4 line. The other line is the UAS_CmOsk line. We use the PiggyBac transposon system to insert the UAS::CmOsk::Tc-eagle mRNA localization signal into *Tribolium* genome. The maternal Gal4 will bind to UAS, which will lead to the maternal expression of *Callosobruchus oskar*. The maternally expressed mRNA will be localized to the anterior of the embryo because the *Tc-eagle* mRNA localization signal. By maintaining and mating these two lines, we are able to examine the initiation of germ plasm assembly in *Tribolium*.

4.2 Materials and Methods

4.2.1 CRISPR target site identification
We choose the first exon of *Tc-eagle* for CRISPR target site screen (Fig. 4.1). The possible target sites were identified by CRISPR Optimal Target Finder (http://tools.flycrispr.molbio.wisc.edu/targetFinder/) (Fig. 4.2). Since we want to nick the double DNA stands, we chose the first two sites for single guide RNA design from these 22 target sites (red box in Fig. 4.2 A). These two sites are 9bp away from each other and are on opposite strands (Fig. 4.2). The PAM for target site 1 is 5’TGG3’, and for target site 2 is 5’CGG3’. The sequences of N18+NGG (PAM) for target site 1 and target site 2 are 5’ CCA GCT GTG CAA GGT CTG (N18) CGG(PAM)3’ and 5’CCG CAG ACC TTG CAC AGC (N18) TGG (PAM)3’, respectively.

### 4.2.2 Single guide RNA production

The single guide RNAs were produced based on the methods of CRISPR on *Drosophila* published by Bassett et al. in 2013 [1]. The forward primer for target site 1 was designed as 5’GAA ATT AAT ACG ACT CAC TAT AGG CCA GCT GTG CAA GGT CTG GTT TTA GAG CTA GAA ATA GC3’ (Lynch lab primer code 783), and the forward primer for target site 2 was designed as 5’GAA ATT AAT ACG ACT CAC TAT AGG CCG CAG ACC TTG CAC AGC GTT TTA GAG CTA GAA ATA GC3’ (Lynch lab primer code 784). The reverse primer for both target sites are the same which was designed as 5’ AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT TAA CTT GCT ATT TCT AGC TCT AAA AC3’ (Lynch lab primer code 582). The primers were submitted to IDT for productions.
These oligonucleotides were used for PCR with Platinum™ PCR SuperMix High Fidelity kit (Thermo Fisher Scientific) without template. The PCR reactions were performed on an Eppendorf thermal cycler with the following program: 35 cycles of (98 °C 10s, 60 °C 30s, 72 °C 15s) and 72 °C 10min, 10 °C ∞. The PCR product was cloned with TOPO® TA Cloning® Kit (Thermo Fisher Scientific), which was transformed into The Mix & Go E. coli competent cells (Zymo Research). Through blue-white screening, white colonies were picked for cell culture, which was used for plasmid extractions. The plasmid containing the DNA fragment for either of the sgRNAs was linearized separately as the templates for sgRNA productions using the HiScribe™ T7 High Yield RNA Synthesis Kit (New England Biolabs). The templates were digested by DNAse I, and the sgRNAs were extracted by phenol chloroform and precipitated by isopropanol. sgRNA pellets were dissolved in nuclease free water (Ambion) and diluted to 1ug/ul. The aliquots were stored at -80 °C.

4.2.3 *In vitro* test of the sgRNAs

The sgRNAs were tested on the genomic DNA fragment that contained the target sites (TS1) in vitro. TS1, 1894 bp long, was amplified from the *Tribolium* genomic DNA using two primers (Lynch lab primer code 798: TGT TTC TAG CAA AGT GTC ATC G, Lynch lab primer code 799: CTA TAA AGC GCA CTT TTA CGA GAT T) with the Platinum™ PCR SuperMix High Fidelity kit (Thermo Fisher Scientific). The PCR program was set as 35 cycles of (94 °C 10s, 50 °C 30s, 72 °C 2 min) and 72 °C 10min, 4 °C ∞. The two CRISPR target sites were about 530 bp away from the start of TS1, which will
give two DNA fragments (~530bp and ~1360bp) if TS1 is cut by the Cas9 nuclease (Fig. 4.3).

The TS1 was then digested with target site 1 sgRNA and target site 2 sgRNA, together and separately. The reaction was set as 30ng target sgRNA 1, 120ng TS1, and 1uL Cas9 nuclease (New England Biolabs, 1uM), and was incubated at 37 °C for one hour. Negative controls that either had no target site 1 and target site sgRNAs or had no Cas9 nuclease, were also performed. The digested products were analyzed by agarose gel electrophoresis.

4.2.4 The donor plasmid productions for CRISPR_Gal4 line and UAS_CmOsk line

The donor plasmid for CRISPR_Gal4 line was to provide Gal4::3Xp3EGFP DNA fragment with flanking arms on each side for HDR DNA repair. The DNA fragment 5'arm::Gal4::3Xp3EGFP::3'arm was constructed in the TA cloning vector. This donor plasmid was ready. The donor plasmid for UAS_CmOsk line will be constructed in the PiggyBac vector to transpose the UAS::CmOsk::mRNA Localization Signal::3xp3RFP into the Tribolium genome. This donor plasmid is not ready so far.

4.2.4 The red flour beetle Tribolium rearing

The beetles were kept at 30 °C incubator with 40-60% humidity in the organic unbleached flour supplemented with yeast powder and Fumagilin (50g yeast and 5g
Fumagilin per 1 kilograms). The flour needed to be kept at -20 °C for two days and then moved to 60 °C for overnight to avoid mites contamination and mixing stains. Then the flour was brought to room temperature before adding yeast and Fumagilin. The eggs take around 4 days to hatch, two weeks in larval stage, and one week in pupa stage at 30 °C, which is around one month for the whole life cycle.

4.2.5 Embryo collection

The beetles, embryos, and flour are separated by going through No. 25 and No.50 sieves in a sequential order. The embryos are transferred on to a basket with the120um nylon mesh and washed several times with DI water to remove excess flour, and then they were dechorionated by placing them in the 50% bleach (Clorox) for 1 minute. The bleaching was repeated one more time. The dechorionated embryos were rinsed with DI water to remove the bleach.

If the embryos are collected for embryo injection, then they are ready to use. If the embryos are collected for in situ hybridization, then the dechorionated embryos are transferred into the fixative (5mL heptane, 2mL PBS and 2mL 10% Methanol free formaldehyde (Polysciences)) for overnight fixation on a shaking platform. After fixation, the aqueous layer was removed, and cold methanol was added to the embryos. The embryos were devitellinized by vigorously shaking cold methanol. After methanol shock, the embryos at the bottom of the vials were collected and stored in Methanol at -20 °C.
4.2.6 Embryonic injection

After one hour egg lay at 25 °C, the embryos were collected and incubated at 25 °C for another hour. The embryos were then dechorionated as described above and aligned vertically on a slide for injection under air.

The injection solution for CRISPR_Gal4 line included target site 1 sgRNA (500ng/uL), target site 2 sgRNA (500ng/uL), Cas9 nickase (500ng/uL), and donor plasmid for HDR DNA repair (200ng/uL), which was loaded into the Femtotips II Microinjection Capillary (Cat. No. 930000043, Eppendorf). The constant pressure was set at 800hpa and the injection pressure was set at 1200hpa to begin with. Since the injection was under air, the needle needed to be cleaned every time after an injection, and the injection pressure also needed to be adjusted periodically as it changed over the course of injection. The injection needed to be done as soon as possible to avoid developmental stage changes since it was conducted at room temperature. After injection, the slide was transferred to an agar plate pre-warmed at 32 °C for about three days incubation till the embryos were hatched. After the first embryo was hatched, the embryos were all moved to the sterilized flour with yeast and Fumagilin to let the larvae develop to adult beetles.

4.3 Results

4.3.1 The CRISPR_Gal4 line design
The direct expression of *Callosobruchus oskar* gene can be toxic to *Tribolium* and might be hard to maintain the transgenic lines. Therefore, we combined the CRISPR/Cas9 and the Gal4/UAS systems along with the PiggyBac transposon system to generate the two transgenic lines.

One of the transgenic lines was named the CRISPR_Gal4 line. Through this line, we wanted to express the *Gal4* gene during oogenesis. Because the maternal expression of *Gal4* would induce the maternal expression of the *Cm-osk* when crossing with the UAS_CmOsk line. In order to achieve this goal, we first thought to find the proper enhancer and promoter for the maternal expression of *Gal4*. However, this process requires a lot of effort and might not be successful. Then we thought we could place the *Gal4* gene under the regulatory elements of a gene that is expressed maternally. In 2005 Bucher et al. showed that *Tc-eagle* is expressed and localized anteriorly in the early embryos [8], which means that *Gal4* can also be expressed maternally if it is placed under the regulatory elements of *Tc-eagle*. In this work, they also claimed that the knockdown of *Tc-eagle* by parental RNA interference (pRNAi) did not drastically interfere with the oogenesis or early development. Although, the knockdown effects on later developmental stages were not characterized, the screen of heterozygous genome editing by CRISPR/Cas9 will avoid the undefined impact on *Tribolium* normal embryogenesis.

When designing this transgenic line, we put the screening process as one of the top priorities. Because the life cycle of the beetle is around one month at 30 °C which is time-consuming. Therefore, we used the 3xP3-EGFP reporter to help us immediately
Fig. 4.4 The 3xP3-EGFP reporter shows the green fluorescence in *Tribolium* eye and nervous system. (A) The EGFP is expressed in the eye and the nervous system of the larva. (B) The EGFP is expressed in the eye and the brain of the pupa. (C) The EGFP is expressed in the adult eyes. This figure is from [6].
Fig. 4.5 CRISPR/Cas9 system for targeting double-strand breaks. This figure is cited from [5]. (A) The two-component system for inducing double-strand breaks. The synthetic guide RNA (sgRNA) contains a region of complementarity to the target site on the DNA, as well as stem loops from the tracrRNA to mediate binding to the Cas9 protein. Cas9 protein is indicated by a yellow circle, cleavage sites by arrowheads, and the protospacer adjacent motif (PAM, NGG) required for cleavage in red. (B) The PCR-based system for generation of sgRNAs. Two oligonucleotides are used to generate the sgRNA template for in vitro transcription (black lines). The first includes a T7 promoter (highlighted in blue) and upstream sequence for efficient in vitro transcription, followed by a GGN_{18} target-site sequence and a portion of the sgRNA stem loops. The second includes the entire sgRNA sequence after the target site. PCR is performed with the two primers but without any other DNA template. This figure is from [5].
identify the individuals that have the *Gal4* gene inserted at the right place. The 3XP3-EGFP reporter was proved to show fluorescence in the eyes and nervous system of *Tribolium* larva and pupa, and in the adult eyes (Fig. 4.4) [6, 9]. This is because the artificial 3xP3 promoter has three binding sites for Pax-6 homodimers, and the pax6 gene was found to be expressed in the eye and brain of *Drosophila* and *Tribolium* [6, 10]. In addition, female beetles can lay many eggs during its life time, which will help us to maintain a stable heterozygous transgenic line.

Given all these considerations, we adopted the CRISPR/Cas9 system to nick the two strands at two CRISPR target sites by Cas9 nickase variant. The reason that we chose to nick both DNA strands was that we can reduce the chance of non-homologous end joining (NHEJ) DNA repair to introduce indels and increase the homology-directed repair (HDR). In order to trigger HDR, we need to provide homologous arms flanking the DNA fragment that was desired to be transferred. For each arm, we cloned a DNA fragment that was around 800 bp long. Therefore, we had the DNA fragment 5’arm::Gal4::3xP3-EGFP::3’arm. This fragment was inserted into the TA cloning vector. The donor plasmid construction was successful.

**4.3.2 Single guide RNA design**

Since we decided to use the first exon of *Tc-eagle* for CRISPR genome editing sites, we identified 21 target sites in total by CRISPR Optimal Target Finder (Fig. 4.1, Fig. 4.2). A target site includes the protospacer sequence (20nt) which is needed for sgRNA
to guide the Cas9 nuclease to this specific site, and the three downstream nucleotides 5'NGG", the protospacer adjacent motif (PAM), which is required for Cas9 nuclease cleavage at the site where is three nucleotides upstream of the PAM.

After examining all the target sites, we chose the first two sites that are on opposite stands and are 9bp apart. The nicking site for target site 1 was between the 10th bp and 11th bp of the first exon, and the nicking site for target site 2 was between the 19th bp and 20th bp of the first exon (Fig. 4.2). The close distance between the two nicking sites would replace only 9 base pairs of DNA fragment, which would potentially minimize the disruptions of the *Tc-eagle* gene that might contain other regulatory elements. The protospacer of target site 1 is 18 nucleotides long followed by its PAM 5'TGG3', and the protospacer of target site 2 is also 18 nucleotides long followed by its PAM 5'CGG3'.

The sgRNAs were designed based on the methods described by Bassett et al. [5]. In this publication, the authors designed unique primers to amplify the template for sgRNA productions. The forward primer is made of the T7 polymerase binding site followed by the gene-specific guide sequence (20nt) and a common sequence overlapping with the last part of the reverse primer (Fig. 4.5). Beside the common part, the reverse primer also had its unique sequence. By PCR with these two primer and no template, the template for sgRNA will be amplified for T7 in vitro RNA production (Fig 4.5). The reverse primer contains all the necessary sequence information for incorporating the sgRNA into the Cas9 nuclease, and the gene-specific guide sequence brings the Cas9-sgRNA complex for specifically targeting.
4.3.3 The UAS_CmOsk line design

The second transgenic line is named as UAS_CmOsk line. Since the expression of genes near the UAS is dependent on the Gal4 expression, the insertion of the Cm-osk can be random in the genome. The PiggyBac transposon system can move the DNA fragment from the PiggyBac vector to the genome, *vice versa*. The PiggyBac transposase that is produced by the helper vector can recognize the inverted terminal repeat sequences (ITRs) on the PiggyBac vector, and then it cuts at the ITRs and inserts the DNA fragment at the TTAA sites in the genome. Therefore, we designed to use the PiggyBac transposon system to insert the DNA fragment UAS::CmOsk::Tc-eagle mRNA Localization signal::3xP3-RFP randomly into the Tribolium genome. The Tc-eagle mRNA Localization signal at the 3’ UTR contains the information to transport and localize the Cm-osk mRNA at the posterior pole of the oocyte. In *Drosophila* it has been proved that the localized oskar mRNA is necessary for germ plasm assembly [7, 11, 12]. The 3xP3-RFP is a reporter that will produce red fluorescent protein in the eye and nervous system when the UAS is bound with Gal4 and the expression is activated.

4.3.4 sgRNAs in vitro test showed that the Cas9-sgRNA complexes can efficiently cut the target sites

Before the *in vivo* injection for CRISPR/Cas9 genome editing, we performed the *in vitro* test on the efficiencies of the sgRNAs. A piece of genomic DNA fragment, the TS1
Fig. 4.6 The genomic DNA fragment (TS1) for the in vitro efficiency test of the two sgRNAs. The forward primer (Lynch lab primer code 798) and the reverse primer (Lynch lab primer code 799) shows the fragment that was cloned for the in vitro test, which is 1894 bp long. The two CRISPR target sites were 521 bp and 530 bp away from the forward primer, respectively.
Fig. 4.7 sgRNAs in vitro test showed that the Cas9/sgRNA complexes can efficiently cut the target sites. The negative controls lane 1 and lane 2 show that the bands were not cut. TS1 DNA fragments in lane 3, lane 4 and lane 5 were cut either separately or together by the Cas9/sgRNA complexes. The bands are the larger fragment which is around 1360pb.
was cloned as the template. The target sites were located around 530bp away from the first base pair (Fig. 4.6). The two sgRNAs were tested separately and together with the Cas9 nuclease to cut the target sites on TS1. To rule out non-specific cutting, we performed two negative controls. The first one was not provided with Cas9 nuclease but was provided with the two sgRNAs, and the second one was not provided with the two sgRNAs but was provided with the Cas9 nuclease (Fig. 4.7). The in vitro test results showed that both sgRNAs, either together or separately, can efficiently guide the Cas9 for specific cutting. Since the two target sites were only 9 bp apart, the larger bands (~1360bp) all looked the same. The smaller bands (~530bp) were not able to show on the gel. This was due to a small amount of genomic DNA template, which was only 120ng per reaction. Therefore, the smaller size of the DNA fragments was hard to identify on the gel. The successful and specific cutting indicated that both sgRNAs were ready for in vivo CRISPR genome editing.

4.3.4 The in vivo injection did not result in any transgenic lines

Thirty eggs were collected after one-hour egg lay at 25 °C and then incubated for another hour. At this time, the embryos were still syncytial, and injection on these embryos would increase the chance for germline transformation. The injection was performed in air instead of in water or in oil to help the wound heal for higher survival rate. The disadvantages of the injection in air was the clog in the delicate glass needle by the yolk. Pressing the clean button to remove the clog or changing the needles more often would solve this problem. Although the normal Tribolium eggs can hatch in the dry environment
of the flour, the injected embryos were placed on the agarose gel for moist till they hatched in order to avoid the embryos being dried out by the flour. Of the 30 embryos, I was only able to gain 12 good injected embryos. After these embryos were hatched, they were transferred to flour to continue the growth.

Since the CRISPR_Gal4 line has the reporter 3XP3-EGFP and the Pax-6 is expressed in the nervous system and the eye since the larva stage, the green fluorescence should be seen if the CRISPR/Cas9 and the HDR DNA repair worked. However, I did not observe the expression of the EGFP in these tissues. This could be resulted from several reasons. First, the CRISPR did not work at all. Even though the \textit{in vitro} test worked, the \textit{in vivo} genome editing would not necessarily work. This is because the genomic DNA fragment in vitro is linear, while the 3D genomic structure could change the landscape of the target sites and might not be easy for Cas9/sgRNA complexes to access. Second, the Cas9/sgRNA complexes did nick both strands, but the HDR was not used but the NHJE DNA repair was. In this case, the reporter along with the Gal4 gene was not inserted into the genome, instead indels might be introduced into the genome.

4.4 Discussion

4.4.1 Does \textit{Tc-eagle} knockout result in disruptions of embryogenesis?

One of the concerns is the \textit{Tc-eagle} knockout would severely disrupt the embryogenesis. In \textit{Drosophila}, eagle is ubiquitously expressed in early embryos, and later it is expressed in some of the neuroblasts [13-15]. Although Bucher et al. showed that the
Tc-eagle knockdown does not drastically interfere with early development or oogenesis, the effects on later stages was not examined [8]. Therefore, the effects of the knockdown should be carefully characterized. Despite the potential knockout effects of Tc-eagle on embryogenesis, heterozygous knockout on one of the two loci could solve this problem. To maintain a stable line, we only need to populate the number of the individuals with the knock-in by mating them with the wild type beetles and screen the transgenes with the marker that show green fluorescence in the eyes and the nervous system.

4.4.2 CRISPR/Cas9 trouble-shooting

In 2015 after this experiment was performed, Gilles et al. successfully applied the CRISPR/Cas9 genome editing to the beetle Tribolium and showed that it introduced the indels [16]. They claimed that 55-80% of the injected individuals were introduced with indels by NHEJ repair, of which 71-100% were able to pass these mutations to the next generations through germline. Comparing to NHEJ repair, the homology-directed knock-in rate was drastically low, with 6% of the injected individuals possessed knocked-in in their germline that can be passed to their offspring.

Considering the 12 embryos I injected that were in good shape, it seemed to make sense that I did not gain any individual that possessed the knock-in DNA. To test if the CRISPR works or not with the amounts of the sgRNAs, Cas9 protein and the donor plasmid I injected in future, it will be necessary to do genotyping to test whether there are indels introduced by NEHJ since it has a higher rate to occur. If not, the reference amount
of the components injected to the embryos can be adjust accordingly. Importantly, injecting a large number of embryos to compensate the low frequency of the HDR that can be transmitted to the next generations is also necessary.

4.4.3 The possible outcomes of the cross between the two transgeneic lines

The reason of designing two transgenic lines was that we were not sure what the direct expression of *Callisobruchus Osk* gene would affect the *Tribolium* embryogenesis. By crossing the two transgenic lines, the Gal4 would bind to USA and activate the expression of Cm-osk during oogenesis. Since the *Cm-osk* mRNA has the *Tc-eagle* mRNA’s 3’UTR, theoretically the *Cm-osk* mRNA would be transported and localized at the anterior pole of the oocyte. Once the *Cm-osk* mRNAs are localized, we expect they are translated into protein, with which the *Cm-osk* mRNAs would form a primitive complex. However, we cannot predict whether this complex can recruit other proteins or mRNAs to initiate the germ plasm assembly or not. So far, we have observed the *Tc-tudor* and *Tc-vasa* are maternally expressed (Personal observation). In *Drosophila*, the *tudor* mRNAs and Vasa protein are recruited to form the germ plasm, and these two genes are essential for ectopic germ plasm assembly [7], which means that once the tudor mRNA and Vasa protein are recruited, other necessary components, if they are available during *Tribolium* oogenesis, would also be assembled.

In this work, we cannot predict whether this will lead to germ cell formation or not. However, we do want to see if the supplement of *Cm-osk* would give any sings of germ
plasm assembly. If this is the case, then it would suggest that the conserved components of germline development by the two modes still exist, and the gain of oskar gene is an innovation to the derived germline specification mode, the maternal provision mode.

4.5 Supplementary materials

4.5.1 Additional file 1: *Callosobruchus oskar* coding sequence

5'ATGGTTTTGGGGACTATCAGAATACTGAAGAAGAAAAATATACAAGATTCTTTT
ATCTATAATATGCCTGTACTGTGAAGACGGCCTTCTTTCTAGACCATGTAGATGAAG
AATTCAAAATATGCGGTGGCTCCATCCCTTTATAAGAAATTTGGAGCGTCATCG
CTTAGATCTTTGGCTAGTTACATTACCAGAAAAATTTATTTATTTGTGCGTGATTGTACAATA
GAGAAGTTTTGATTCAAGCAAGTAAAAATCAATGCATATTAAAGAGATTTTAA
AGCAAAAGAAATGTGCTGGAAAAAGTGCCACATGTAGAAATGACTTTCGAAACTAC
AACTTTTATCAAAGAAAAAAGGATATCATCAAAGACCACCTCCGGAAGGATCCTTT
GTACATGCATCAATCAAATATCCTGCAATTATGATTATAGAAATGTTTAGTTACATTCT
GCAAATTGGATATGTATGCTGCCATTGTTTTACAAACATCCAGGCTTTTAGAGATGAC
TTTTTCTTAGATATAGGCACAAAATAAATGATTAGGATATTATGTACCTGATACTAGGTATC
AAGCAATGTGGCTCTATGTGATCTGGACAAACAAATAGCCTCGTGGCAAAAGGT
GAAAAGAATCAGAAAATGAGCTCCAAAGGGTGCATTGTGATGATGGGATTTCAAGATTT
GCTAATGGGCAAATACAAACAAACATGATTTATGATTGATTACTGAC
ATTGAAGAATAAAGCAGCAGTTACCCTTGTCAACAATTTCCACGCTCTCCAAAGC
TGCCCTAACTCCAGAAAAATTGCTGCAGGCTCTGATTTATAATCAAGGCAATTATGG
ACTATTCTTGTTCTTCTGACTTTAAATGCAATGTAATTGACATGCATAGAATATTCCA
TCAGGAGCAATCTACTTTTCAGAAGAGACCTTTGACAGGCTCTCAAAAGTGGCCAAAA
ACGATCAGTATAAGGTCTTCAGTGATTATGGAAGGAAGATATTCCCTCAGGACACTTTAA
AGACTTGGCCTGAAAGAAGCATTAGAAGCTGGACATTGA3’

4.6 Literature Cited


15. Dittrich R, Bossing T, Gould AP, Technau GM, Urban J: The differentiation of the serotonergic neurons in the Drosophila ventral nerve cord depends on the

CHAPTER 5

DISCUSSION AND OUTLOOK
5.1 New tools for deep functional analysis on the new genes we found

The RNA-seq results on *Nasonia* oosome gave us a lot of novel genes that might be directly involved in the germline development (Chapter 2). This also suggests that we can extend this method to other insects and expect to find genes like these too. The comprehensive analyses on the insect germline specification will shed light on how the germline specification mechanism evolved over the course of evolution. However, with the unavailable access to the convenient and high-through output knockdown or knockout techniques, we are limited to the preliminary functional studies on those novel genes that we found. New techniques on deep functional study need to be developed. For example, the CRISPR/Cas9 is a good way to establish stable transgenic lines. In addition, I also tried to develop a live imaging tool to observe the knockdown effects on the embryogenesis, in which I injected the *in vitro* synthesized mRNAs for membrane and nucleus markers and the dsRNA for the specific gene that I wanted to knock down. The problem I had was that the mRNA translation efficiency was too low to gain high resolution images for functional analysis. Troubleshooting on improving the translation efficiency should be done in future.

5.2 Identification of the protein composition in *Nasonia* oosome

The major components of the germ plasm are mRNAs and proteins. RNA components were revealed during my Ph.D. study. However, little is known about the proteins localized in the germ plasm. It is already known that the proteins localized in the *Nasonia* oosome and Drosophila polar granules play important roles in germline
development, such as Vasa and Tudor, etc. [1-4]. Therefore, identifying the protein composition by mass spectrum will let us understand how the germ plasm is organized, and what other genes contribute to the germline development.

5.3 Single cell sequencing on *Nasonia* pole cells, migrating germ cells, and *Tribolium* PGCs

The maternally inherited germ plasm determines the germ cell fate. We used RNA-seq to analyze the mRNAs that partially contribute to the germ cell fate determination. After the zygotic transcriptome is activated, we are not sure what genes are expressed in the PGCs. The single cell sequencing technique will let us know the RNA expression landscape in the germ cells. It would be intriguing to perform this technique on the PGCs from *Tribolium*, which are specified by the zygotic inductive signals [5]. We would expect to find the initial genes that are turned on to induce extraembryonic stem cells to become PGCs.

5.4 Overcome the technical limitations on dissecting the germ plasm assembly mechanism in *Callosobruchus*

The effects of knockdown of the genes on germ plasm assembly can be examined in the early *Callosobruchus* embryos. *Vasa* and *tudor* mRNAs are localized in germ plasm in early embryos (Fig. 3.1), and could be used as the germline marker. However, collecting and processing the embryos for *in situ* hybridization and immunostaining is still
challenging, which is due to the thick eggshell, and cloudy and sticky embryos. I tried to use hand dissect the eggshells off the fixed embryos, but it did not work well. It was time-consuming and easy to damage the embryos. Therefore, sectioning the embryos into slices seems to be a feasible way to prepare the samples for in situ hybridization and immunostaining. Similar experiments had successfully done before [6]. Adapting this protocol to Callosobruchus should not be hard. On the other hand, we are still interested in examining the germ plasm in the oocytes, for which we do not have a marker. I purified the Oskar protein to make the polyclonal antibody, which was a success. Other antibodies can also be made using the same way I did for Oskar, such as Vasa and Tudor.

In addition, we found that bruno mRNA is localized in the anterior pole of the oocyte and the early embryos, which suggests a potential role in anterior-posterior patterning. Indeed, the knockdown results suggested a potential role for bruno (Fig. 3.5). Therefore, finding other anterior and posterior markers would enhance our understanding of the roles of bruno in germline development and anterior-posterior patterning.

5.5 Literature Cited


APPENDIX
APENDIX A: Permission to reuse published materials

This Agreement between Honghu Quan ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>4450841161869</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Oct 16, 2018</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>John Wiley and Sons</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Evolution &amp; Development</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Anterior localization of maternal mRNAs in a short germ insect lacking bicoid</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Gregor Bucher, Laila Farzana, Susan J. Brown, et al</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Feb 25, 2005</td>
</tr>
<tr>
<td>Licensed Content Volume</td>
<td>7</td>
</tr>
<tr>
<td>Licensed Content Issue</td>
<td>2</td>
</tr>
<tr>
<td>Licensed Content Pages</td>
<td>8</td>
</tr>
<tr>
<td>Type of use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>University/Academic</td>
</tr>
<tr>
<td>Format</td>
<td>Print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>Figure/table</td>
</tr>
<tr>
<td>Number of figures/tables</td>
<td>1</td>
</tr>
<tr>
<td>Original Wiley figure/table number(s)</td>
<td>Figure 3</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Evolution of insect germline specification</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Dec 2018</td>
</tr>
<tr>
<td>Expected size (number of pages)</td>
<td>200</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>Honghu Quan</td>
</tr>
<tr>
<td></td>
<td>900 S Ashland Ave,</td>
</tr>
<tr>
<td></td>
<td>MBBR room 4018</td>
</tr>
<tr>
<td></td>
<td>M/C 567</td>
</tr>
<tr>
<td></td>
<td>CHICAGO, IL 60607</td>
</tr>
<tr>
<td></td>
<td>United States</td>
</tr>
<tr>
<td></td>
<td>Attn: Honghu Quan</td>
</tr>
<tr>
<td>Publisher Tax ID</td>
<td>EU826007151</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Terms and Conditions</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX A (continued)

TERMS AND CONDITIONS
This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., "CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, and any CONTENT (PDF or image file) purchased as part of your order, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts, you may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the
continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensor and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or
APPENDIX A (continued)

excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY’s prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC’s Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties’ successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC’s Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state’s conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-
APPENDIX A (continued)

Creative Commons Attribution Non-Commercial License
The Creative Commons Attribution Non-Commercial (CC-BY-NC) License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes. (see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License
The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations
Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Further details can be found on Wiley Online Library http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
APPENDIX A (continued)

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Oct 19, 2018

This Agreement between Honghu Quan ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>4450850183678</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Oct 16, 2018</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>Springer Nature</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Nature Reviews Genetics</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Innovations: Applications of insect transgenesis</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Ernst A. Wimmer</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Mar 1, 2003</td>
</tr>
<tr>
<td>Licensed Content Volume</td>
<td>4</td>
</tr>
<tr>
<td>Licensed Content Issue</td>
<td>3</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/university or research institute</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>High-res required</td>
<td>no</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>no</td>
</tr>
<tr>
<td>Circulation/distribution</td>
<td>&lt;501</td>
</tr>
<tr>
<td>Author of this Springer Nature content</td>
<td>no</td>
</tr>
<tr>
<td>Title</td>
<td>Evolution of insect germline specification</td>
</tr>
<tr>
<td>Institution name</td>
<td>University of Illinois at Chicago</td>
</tr>
<tr>
<td>Expected presentation date</td>
<td>Dec 2018</td>
</tr>
<tr>
<td>Portions</td>
<td>Figure 1 g-i</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>Honghu Quan</td>
</tr>
<tr>
<td></td>
<td>900 S Ashland Ave,</td>
</tr>
<tr>
<td></td>
<td>MBRB room 4018</td>
</tr>
<tr>
<td></td>
<td>M/C 567</td>
</tr>
<tr>
<td></td>
<td>CHICAGO, IL 60607</td>
</tr>
<tr>
<td></td>
<td>United States</td>
</tr>
<tr>
<td></td>
<td>Attn: Honghu Quan</td>
</tr>
<tr>
<td>Billing Type</td>
<td>Invoice</td>
</tr>
<tr>
<td>Billing Address</td>
<td>Honghu Quan</td>
</tr>
<tr>
<td></td>
<td>900 S Ashland Ave,</td>
</tr>
</tbody>
</table>
Terms and Conditions

Springer Nature Terms and Conditions for RightsLink Permissions

Springer Nature Customer Service Centre GmbH (the Licensor) hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where print only permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where 'reuse in a dissertation/thesis' has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the Journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this
APPENDIX A (continued)

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor’s approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)

For Advance Online Publication papers:
Reprinted by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication), advance online publication, day month year (doi: 10.1038/sj.JOURNAL ACRONYM].

For Adaptations/Translations:
Adapted/Translated by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [the Licensor]: on behalf of Cancer Research UK: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)

For Advance Online Publication papers:
Reprinted by permission from The [the Licensor]: on behalf of Cancer Research UK: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication), advance online publication, day month year (doi: 10.1038/sj. JOURNAL ACRONYM]

For Book content:
Reprinted/adapted by permission from [the Licensor]: [Book Publisher (e.g. Palgrave Macmillan, Springer etc) [Book Title] by [Book author(s)] [COPYRIGHT] (year of publication)

Other Conditions:

Version 1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
This Agreement between Honghu Quan ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>4463190037294</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Nov 06, 2018</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>Springer Nature</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Springer eBook</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Drosophila melanogaster Oogenesis: An Overview</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>John M. McLaughlin, Diana P. Bratu</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Jan 1, 2015</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/university or research institute</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>no</td>
</tr>
<tr>
<td>Circulation/distribution</td>
<td>&lt;501</td>
</tr>
<tr>
<td>Author of this Springer Nature content</td>
<td>no</td>
</tr>
<tr>
<td>Title</td>
<td>Evolution of insect germline specification</td>
</tr>
<tr>
<td>Institution name</td>
<td>University of Illinois at Chicago</td>
</tr>
<tr>
<td>Expected presentation date</td>
<td>Dec 2018</td>
</tr>
<tr>
<td>Portions</td>
<td>Fig. 1 on page 2</td>
</tr>
</tbody>
</table>
| Requestor Location    | Honghu Quan  
900 S Ashland Ave,  
MBRB room 4018  
M/C 567  
CHICAGO, IL 60607  
United States  
Attn: Honghu Quan |
| Billing Type          | Invoice |
| Billing Address       | Honghu Quan  
900 S Ashland Ave,  
MBRB room 4018 |
M/C 567
CHICAGO, IL 60607
United States
Attn: Honghu Quan

Total
0.00 USD

Terms and Conditions

Springer Nature Terms and Conditions for RightsLink Permissions

Springer Nature Customer Service Centre GmbH (the Licensor) hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where print only permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where 'reuse in a dissertation/thesis' has been selected the following terms apply: Print rights of the final author’s accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the Journal/book’s homepage. Our required acknowledgement format is in the Appendix below.
APPENDIX A (continued)

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)

For Advance Online Publication papers:
Reprinted by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:
Reprinted/adapted/translated by permission from [the Licensor]: on behalf of Cancer Research UK: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)

For Advance Online Publication papers:
Reprinted by permission from The [the Licensor]: on behalf of Cancer Research UK: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [the Licensor]: [Book Publisher (e.g. Palgrave Macmillan, Springer etc)] [Book Title] by [Book author(s)] [COPYRIGHT] (year of publication)

Other Conditions:
APPENDIX A (continued)

Version 1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
VITA
Honghu Quan
312-395-0877  900 S Ashland Ave, Chicago, IL 60607  hquan2@uic.edu

EDUCATION

• **University of Illinois at Chicago**, Chicago, IL
  Ph.D. Student in Biological Sciences, August 2012 – Present
  Expected Thesis Defense: October 2018

• **Illinois Institute of Technology**, Chicago, IL
  Master Student in Biophysics, August 2011 – May 2012

• **Huazhong Agricultural University**, Wuhan, China
  Bachelor of Veterinary Medicine, June 2011

RESEARCH EXPERIENCE

**Ph.D. Student** in Dr. Jeremy Lynch’s Laboratory (January 2013-Present)

• Identified novel genes in the wasp *Nasonia* germline development using RNAseq and characterized their functions by embryonic RNA interference

• Using the CRISPR/Cas9 and GAL4/UAS systems to knock in the bean beetle *Callosobruchus oskar* gene into the genome of the red flour beetle *Tribolium*

• Purified the *Callosobruchus* Oskar protein and successfully made the polyclonal antibody

• Developed workflow for RNA interference on *Nasonia* syncytial embryos and workflow for fluorescent *in situ* hybridization on *Callosobruchus* ovaries

**Visiting Researcher** in Dr. Michael Eisen’s Laboratory at UC Berkeley (December 2013)

• Learned how to cryo-section single *Tribolium* early embryo into at least ten slices and extract mRNA from each slide for RNAseq on the EDEN award

**Rotation Student** in Dr. Hua Jin’s Laboratory (November - December 2012)

• Constructed plasmid to quantitatively measure the cAMP level in primary cilia of neurons by florescent resonance energy transfer (FRET)

**Rotation Student** in Dr. Brain Kay’s Laboratory (September - November 2012)

• Successfully selected monobodies from FN3 phage library against CENTB2 protein by phage display

**Research Volunteer** in Dr. Qun-Tian Wang’s Laboratory at UIC (August 2011 - May 2012)

• Purified different domains of Asxl2 protein and pulled down proteins interacting with these domains

CORE TECHNICAL SKILLS
• **DNA Biology**: PCR, RT-PCR, qPCR analysis, Mutagenesis, Cloning, CRISPR/Cas9 genome editing

• **RNA biology**: *In situ* hybridization, Library preparation, RNA sequencing, RNA interference, mRNA-based live imaging

• **Protein Biology**: Immunohistochemistry analysis, Immunocytochemistry Assay, Phage Display, Protein purification, Fluorescence Resonance Energy Transfer (FRET), Immunoprecipitation, Mass spectrum (MS), ELISA, Western blotting, SDS-PAGE

• **Cell Biology**: Mammalian tissue culture, cell culture, transfection

• **Organism Model**: Mouse, Rabbit, *Nasonia, Drosophila, E. coli*, Embryonic microinjection, Histological sectioning, tissue processing

• **Bioinformatics**: Linux commands, RNAseq analysis, Python

• **Imaging**: Confocal Microscopy, Scanning Electron Microscopy, Imaris, Photoshop, ImageJ

**SUPERVISION EXPERIENCE**

Supervision of master student (January 2014-May 2015)
- Identifying *Tribolium castaneum* maternal enhancers to drive the ectopic expression of exogenous genes

Supervision of high school student (October 2017-May 2018)

Teaching Assistant (August 2012-Present)
- Taught Cell Biology lab course, Comparative Vertebrate Anatomy lab course (Sharks, Cats, Rabbit, Mouse), Genetics lab course
- Leading discussions in Genetics lecture course

**FELLOWSHIPS AND AWARDS**
- Evo-Devo-Eco Network (EDEN) Research Exchange Funds (December 2013)

**CONFERENCE PRESENTATIONS**
- 2013 Midwest Drosophila Conference, poster presentation
- 2014 Midwest Drosophila Conference, poster presentation
- 2015 Pan-American Society for Evolutionary Developmental Biology Inaugural Meeting
- The Allied Genetics Conference, poster presentation (2016)
- 2017 ASBMB Special Symposia Series: Evolution and Core Processes in Gene Expression
