



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO
DI INGEGNERIA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

**CORSO DI LAUREA MAGISTRALE IN
BIOINGEGNERIA**

Modelli e analisi di dati biomedicali

**“ANALISI DELL'ESPRESSIONE GENICA DIFFERENZIALE
DELLE CELLULE SPERMATICHE IN VIA DI SVILUPPO DI
ANOPHELES GAMBIAE, PRINCIPALE VETTORE DELLA MALARIA”**

Relatore: Prof. Enrico Lavezzo

Laureanda: Maria Pia Scarfone

ANNO ACCADEMICO 2023 – 2024

Data di laurea: 27 febbraio 2024

Abstract

La malaria è una malattia infettiva trasmessa attraverso la puntura di zanzare femmine infette del genere *Anopheles*. La lotta contro la malaria include l'utilizzo di strategie che prevedono il controllo del vettore e la somministrazione di farmaci e vaccini per ridurre il numero di casi e decessi. L'eliminazione a livello locale, però, non è più sufficiente. Per contrastare completamente la malattia è necessario eradicarla. L'eradicazione prevede l'eliminazione definitiva della malaria a livello globale e necessita di tecniche scientifiche avanzate che pongono l'attenzione sulla genetica del vettore. In questo lavoro di tesi viene eseguita l'analisi bioinformatica di trascritti di RNA provenienti da quattro linee cellulari estratte dalle gonadi maschili di *A. gambiae*: premeiotiche, meiotiche (spermatociti primari e secondari) e postmeiotiche. Per prima cosa è stato effettuato un controllo di qualità dei dati, seguito da una fase di filtraggio delle reads di sequenziamento. Successivamente è stato eseguito l'allineamento al genoma di riferimento, i cui risultati sono stati poi visualizzati e analizzati. La quantificazione dei trascritti è stata eseguita in due diverse condizioni, ovvero presenza e assenza del genoma mitocondriale. L'ultimo step dell'analisi bioinformatica è stato lo studio dell'espressione genica differenziale. Lo studio dell'espressione genica differenziale permette di chiarire quali geni sono coinvolti in punti temporali specifici durante il processo di spermatogenesi. L'identificazione di particolari fattori genetici coinvolti nella fertilità, vitalità e riproduzione delle zanzare può risultare d'aiuto nell'implementazione di nuove strategie di gene drive. Ad oggi, il rilascio di vettori transgenici è uno strumento utile al raggiungimento dell'obiettivo di eliminazione della malaria, mediante la soppressione delle popolazioni di zanzare selvatiche o la riduzione della loro capacità di trasmettere agenti patogeni.

Indice

Capitolo 1	8
La Malaria.....	8
1.1 Diffusione geografica della malaria e trasmissione	8
1.2 Sintomi.....	10
1.3 Eziologia	11
1.4 Ciclo di vita del <i>Plasmodium</i>	12
1.5 Il vettore della malaria: <i>Anopheles gambiae</i>	13
1.5.1 Ciclo di vita e morfologia delle zanzare <i>Anopheles gambiae</i>	13
Capitolo 2	15
Prevenzione, strategie di contenimento e trattamento della malaria	15
2.1 Le strategie di controllo del vettore	15
2.1.1 Le zanzariere da letto impregnate con insetticidi.....	15
2.1.2 Gli insetticidi ad azione residua nelle abitazioni.....	16
2.1.3 Il controllo larvale	17
2.2 La chemioterapia preventiva.....	18
2.3 I vaccini.....	19
2.3.1 I vaccini pre-eritrocitari.....	20
2.3.1.1 Il vaccino RTS,S/AS01	20
2.3.1.2 Il vaccino R21/MM	21
2.3.2 I vaccini eritrocitari	22
2.3.3 I vaccini bloccanti la trasmissione	23
2.4 Lo sviluppo di una popolazione resistente.....	23
2.4.1 Resistenza ai farmaci.....	23
2.4.2 Resistenza agli insetticidi	24
Capitolo 3	25
Il ruolo dell'ingegneria genetica nel controllo della malaria.....	25
3.1 Il controllo genico	25
3.2 Gene drive.....	27
3.2.1 Requisiti del gene drive.....	29
3.2.2 Impatto epidemiologico di un sistema gene drive.....	29
3.3 Sistemi gene drive.....	31

3.3.1	Gi elementi trasponibili	31
3.3.2	I microrganismi ereditari	32
3.3.3	La sottodominanza genetica	33
3.3.4	L'arresto embrionale dominante con effetto materno (MEDEA).....	33
3.3.5	Il drive meiotico	34
3.3.6	I geni dell'endonucleasi homing (HEG)	35
3.3.7	Gene drive basati su CRISPR/Cas9 (CCGD).....	37
3.4	Limiti e sfide del gene drive	40
3.5	Biosicurezza dei sistemi gene drive	41
3.6	Regolamentazione e normativa dei sistemi gene drive.....	43
Capitolo 4		45
Applicazione di strategie di controllo genetico in letteratura.....		45
4.1	Meccanismi di determinazione sessuale in <i>D.melanogaster</i> e <i>A. gambiae</i>	45
4.2	<i>Doublesex</i> e <i>Yob</i> in <i>A. gambiae</i>	46
4.3	Studi	47
4.3.1	Gene drive basato su CRISPR/Cas9 mirato a <i>dsx</i>	47
4.3.2	Gene drive mirato a <i>Yob</i>	48
4.3.3	Elementi regolatori della linea germinale	49
Capitolo 5		51
Materiali e Metodi		51
5.1	Dataset	51
5.2	Analisi bioinformatica.....	51
5.2.1	Controllo di qualità delle reads	52
5.2.2	Processamento delle reads.....	53
5.2.3	Allineamento al genoma di riferimento.....	55
5.2.4	Quantificazione dei trascritti	55
5.2.5	Analisi dell'espressione genica	57
5.2.6	Analisi dell'espressione genica differenziale	57
5.2.7	Analisi dell'arricchimento funzionale	59
Capitolo 6		62
Risultati.....		62
6.1	Quality Control	62
6.2	Allineamento al genoma di riferimento	64

6.3 Quantificazione dei trascritti.....	67
6.4 Espressione Genica prima della rimozione dei geni mitocondriali	69
6.5 Espressione genica dopo la rimozione dei geni mitocondriali	71
6.6 Espressione Genica differenziale.....	73
6.7 Arricchimento funzionale	75
Capitolo 7	82
Discussione.....	82
7.1 Quality control	82
7.2 Allineamento al genoma di riferimento	84
7.3 Quantificazione dei trascritti.....	84
7.4 Espressione genica prima della rimozione dei geni mitocondriali	86
7.5 Espressione genica dopo la rimozione dei geni mitocondriali	88
7.6 Espressione genica differenziale.....	88
7.7 Arricchimento funzionale	88
Bibliografia.....	91

Capitolo 1

La Malaria

La malaria è una malattia infettiva, causata da parassiti appartenenti al genere plasmodio, i quali vengono trasmessi all'uomo attraverso le punture di zanzare infette del genere *Anopheles*. Il *Rapporto mondiale sulla malaria 2023* fornito dall'Organizzazione Mondiale della Sanità segnala che a livello globale, nel 2022, si sono verificati 249 milioni di casi e 608.000 decessi in 85 Paesi endemici [1]. I report pubblicati annualmente permettono di monitorare tutte le strategie applicate in termini di prevenzione, diagnosi, trattamento, sorveglianza ed eliminazione. La raccolta, l'analisi e l'interpretazione dei dati è fondamentale per la pianificazione di interventi sanitari mirati al controllo e all'eliminazione della malaria. A tal proposito nel 2015 è stata definita *La Strategia tecnica globale dell'OMS per la malaria 2016-2030* che prevede la riduzione dell'incidenza globale dei casi e dei tassi di mortalità del 90% e l'eliminazione della malaria in almeno 35 Paesi entro il 2030 [2].

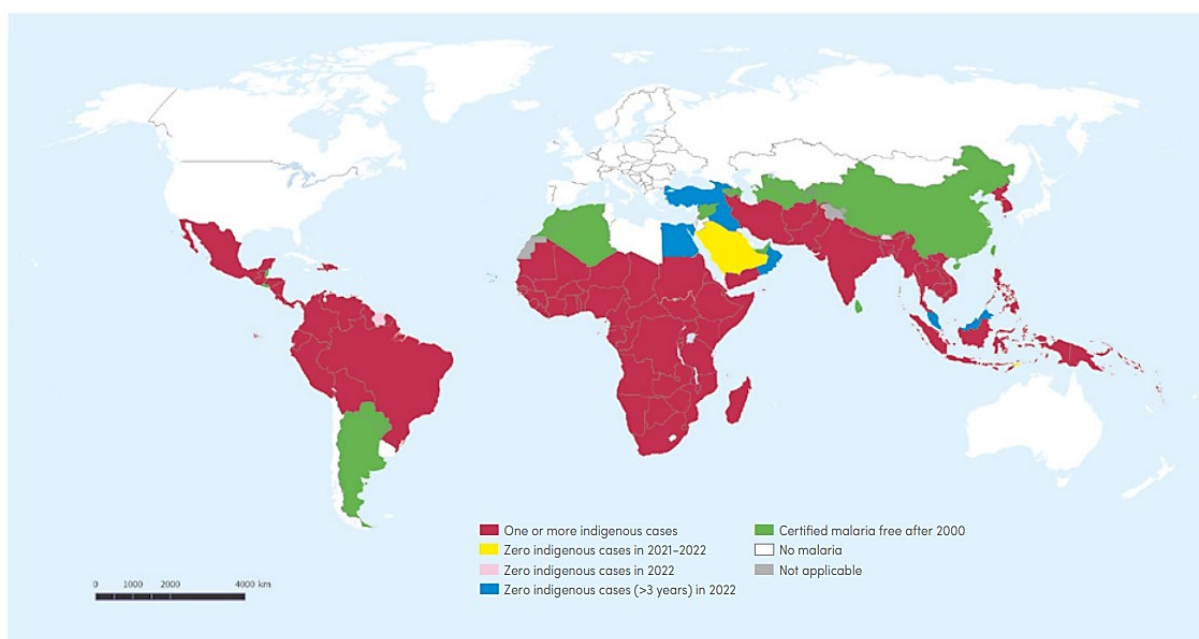


Figura 1. Casi di malaria nel 2022 registrati per Paese [3].

1.1 Diffusione geografica della malaria e trasmissione

La malaria è diffusa principalmente nelle aree tropicali e sub tropicali (Figura 1). In Africa è altamente endemica, in particolare, nell'Africa sub-sahariana dove si registra il maggior numero

di contagi e decessi. Nel 2021, secondo i dati pubblicati dall'Organizzazione mondiale della Sanità, il continente africano ha registrato il 95% dei casi totali stimati. I Paesi più colpiti sono stati Nigeria, Repubblica Democratica del Congo, Uganda e Mozambico [3].

Nei Paesi del Sud-est asiatico, Medio Oriente, Asia Centrale, Pacifico e America Centrale e Meridionale è diffusa con diversi gradi di endemicità.

In Europa, invece, i dati pubblicati a giugno 2023 dal Sistema di sorveglianza TESSy, e relativi all'anno 2021, hanno registrato un numero di casi pari a 4.855. Anche l'Italia, nonostante sia stata dichiarata libera dalla malaria dal 1970, segnala annualmente qualche centinaio di casi (Figura 2).

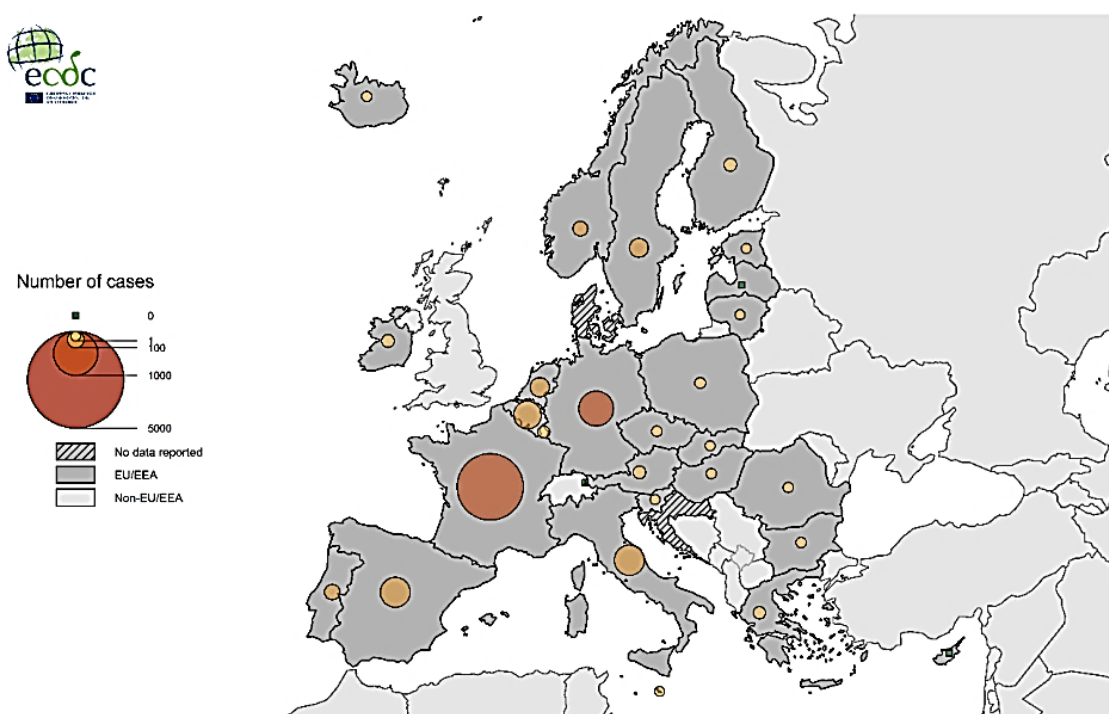


Figura 2. Casi di malaria confermati in Europa nel 2021 [2].

La trasmissione della malaria umana è influenzata da fattori relativi all'ambiente, al clima, alle capacità vettoriali delle zanzare, al parassita e all'ospite.

L'ambiente influenza l'interazione tra l'uomo e le zanzare che mostrano una propensione maggiore a pungere durante la notte in ambienti chiusi. Questa propensione è definita endofagia.

La temperatura, le precipitazioni e l'umidità sono i principali fattori climatici che condizionano la diffusione e la velocità di trasmissione della malaria, agendo sulla sopravvivenza e longevità delle zanzare e sulla velocità di riproduzione del parassita nel vettore. Una temperatura troppo

fredda o calda dell'acqua può determinare un minor numero di uova schiuse o l'alterazione della durata di vita della zanzara di circa una settimana, se si verifica una variazione anche solo di 1°C rispetto alla temperatura ottimale compresa tra i 18°C e i 26°C. Le basse temperature riducono la diffusione della malaria, di conseguenza, ad altitudini superiori ai 1.800-2.000 metri sul livello del mare la malattia non è più endemica, nonostante la presenza di *Anopheles gambiae* sia stata confermata anche ad altitudini di 3.000 metri. Il vento, la durata della luce del giorno e il ritmo circadiano alterano aspetti nel comportamento delle zanzare quali ad esempio l'alimentazione, il riposo e la deposizione delle uova [4].

La capacità vettoriale, ovvero l'abilità e l'efficienza delle zanzare di trasmettere un patogeno ad un ospite suscettibile, è influenzata dalla sopravvivenza e densità del vettore, dallo sviluppo del Plasmodio, il cui ciclo vitale è indipendente dalla disponibilità di acqua e umidità, e dalla frequenza dei pasti di sangue [4]. In relazione al numero di pasti ematici giornalieri (a) e alla proporzione (p) di anofeline che sopravvivono un giorno, è stata realizzata una classifica che permette di definire l'endemicità della malaria. Il valore dell'indice di stabilità $s = \frac{a}{-\ln(p)}$ permette di distinguere la malattia in stabile (indice superiore a 2.5), instabile (indice inferiore a 0.5) e a stabilità intermedia (indice compreso tra 0.5 e 2.5) [5].

Un'elevata trasmissione della malaria tra la popolazione porta allo sviluppo di un'immunità adattativa al plasmodio, riducendo il rischio di contrarre la malattia in forma grave. I bambini di età inferiore a 5 anni e le donne incinte registrano un elevato numero di casi e decessi a causa di una maggiore suscettibilità alla malattia. Solo in caso di sopravvivenza, si sviluppa un'immunità adattativa nei bambini e un'immunità parziale nelle donne incinte. In questo caso, la malaria incide gravemente sullo sviluppo del feto, determinando un basso peso alla nascita e nei casi più gravi, invece, porta all'aborto precoce e alla morte della madre e/o del nascituro. Altri gruppi di popolazione a rischio sono rappresentati dalle persone affette da HIV/IDS e da persone che non presentano immunità al plasmodio. Lo spostamento da aree ad elevata endemicità ad aree a bassa endemicità, in assenza di immunità, potrebbe comportare lo sviluppo di gravi epidemie [6].

1.2 Sintomi

I sintomi malarici compaiono solitamente entro 7-30 giorni dalla puntura della zanzara infetta e possono essere diversi in relazione al plasmodio coinvolto, alla condizione del paziente e all'età. I primi sintomi compaiono alla fine del primo ciclo di riproduzione che avviene nei globuli rossi, i quali rompendosi rilasciano i parassiti. Si manifestano brividi, nausea, affaticamento, cefalea, dolori diffusi e febbre che, in una prima fase, insorge in maniera

imprevedibile, successivamente diventa periodica presentandosi ad intervalli regolari e raggiungendo i 41°C.

Si tratta di sintomi lievi che possono essere difficili da attribuire alla malaria in quanto simili ad altre malattie di gravità minore. I sintomi più gravi includono:

- anemia grave: riduzione del numero di globuli rossi, dovuta alla lisi di quelli infetti e alla rapida senescenza di quelli non infetti eliminati nella milza;
- milza ingrossata (splenomegalia): a causa della presenza di molti parassiti. Può presentarsi tumefatta e subire rotture;
- malaria cerebrale: determina perdita di conoscenza, convulsioni, edema cerebrale, coma e morte. La causa è l'adesione dei globuli rossi infetti alle pareti dei piccoli vasi dei vari organi, determinando occlusioni e, di conseguenza, ischemie.

Altri sintomi sono insufficienza renale, sofferenza respiratoria, ittero, acidosi metabolica, ipoglicemia.

A seconda dei sintomi sviluppati dal paziente è possibile fare una distinzione tra malaria non complicata e malaria severa. La malaria non complicata, asintomatica o facilmente curabile, si presenta principalmente nei soggetti che hanno sviluppato un'immunità nelle aree endemiche. Al contrario, la malaria severa colpisce principalmente bambini e donne in gravidanza [7] [8] [9].

È possibile fare un'ulteriore classifica dell'endemicità malarica, in relazione alla percentuale di bambini esaminati, di età compresa tra i 2 e i 9 anni, che presentano splenomegalia. Se l'ingrossamento palpabile della milza è presente in più del 75% dei casi esaminati si parla di malaria oleoendemica, tra il 51% e 75% si definisce malaria iperendemica, tra l'11% e il 50% è mesoendemica e al di sotto del 10% è detta ipoendemica [4].

1.3 Eziologia

La malaria è una malattia infettiva causata da protozoi parassiti del genere *Plasmodium* e si trasmette attraverso la puntura di zanzare femmine infette definite "vettori della malaria". Esistono varie specie di *Plasmodium* in grado di infettare diversi gruppi di vertebrati, inclusi gli esseri umani. Le specie di *Plasmodium* specifiche per l'uomo responsabili della malaria umana sono:

- *Plasmodium falciparum*: responsabile della febbre terzana¹ maligna, principalmente diffuso nell'Africa sub-Sahariana, Asia, America Latina, Papua-Nuova Guinea e isole del Pacifico Orientale;
- *Plasmodium malariae*: responsabile della febbre quartana benigna e diffuso nelle aree tropicali e sub-tropicali;
- *Plasmodium ovale*: che provoca la febbre terzana benigna ed è comune in Africa Occidentale;
- *Plasmodium vivax*: che provoca la febbre terzana benigna diffusa in Africa Orientale, Asia, America Latina e zone a clima temperato [6] [10];

1.4 Ciclo di vita del *Plasmodium*

La malaria è trasmessa principalmente attraverso la puntura di zanzare femmine infette appartenenti al genere *Anopheles*.

L'infezione malarica inizia quando una zanzara *Anopheles* femmina, durante un pasto ematico effettuato su una persona affetta da malaria, ingerisce sangue contenente le cellule riproduttive (gametociti) del parassita. La specie del plasmodio ingerito e la temperatura influenzano la durata del ciclo di sviluppo del parassita che avviene nel corpo della zanzara. A seguito della riproduzione e dello sviluppo, gli sporozoiti (forme infettanti del plasmodio) migrano verso le ghiandole salivari della zanzara pronti per essere iniettati in un organismo ospite al momento della puntura. Nel soggetto infetto, i parassiti sfuggono all'azione del sistema immunitario e si dirigono verso il fegato invadendo gli epatociti, all'interno dei quali avviene la prima fase di moltiplicazione che culmina con la generazione di uno schizonte e la lisi delle cellule epatiche. La rottura degli epatociti determina il rilascio di merozoiti mononucleati che infettano gli eritrociti, all'interno dei quali avviene il secondo ciclo di riproduzione, producendo generazioni di parassiti ogni 48 ore (terzana) o 72 ore (quartana). I gametociti prodotti rimangono nel circolo sanguigno per settimane cosicché altre zanzare possano infettarsi (Figura 3) [10] [8].

La trasmissione di plasmodi può verificarsi anche tramite contagio ematico diretto, ad esempio nel caso di trasfusione di sangue contaminato, trapianto d'organo contaminato o, ancora, tramite l'uso di un ago già utilizzato da una persona affetta da malaria. Raramente si verifica la trasmissione del plasmodio durante la gravidanza, da madre infetta a feto tramite la placenta (trasmissione transplacentare) [6].

¹ I termini di terzana e quartana si riferiscono al ciclo di replicazione del parassita nel sangue umano e alla manifestazione di febbre intermittente, ogni 48 ore (terzana) o ogni 72 ore (quartana).

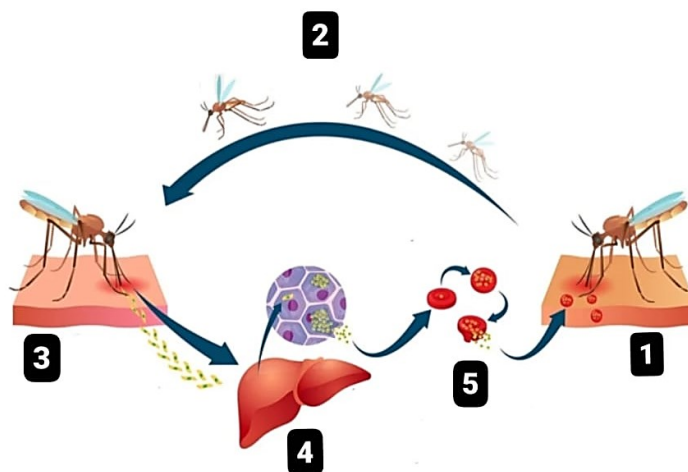


Figura 3. Ciclo di vita dei Plasmodi: (1) la zanzara ingerisce i plasmodi tramite un pasto ematico. (2) Il parassita completa il suo ciclo di sviluppo nel corpo dell'insetto generando sporozoiti che migrano nelle ghiandole salivari e (3) vengono inoculati nell'ospite. (4) Nel fegato vengono generati merozoiti che infettano gli eritrociti a seguito della rottura degli epatociti. (5) Gli eritrociti scoppiano e liberano ulteriori gametociti che rimangono nel circolo sanguigno.

1.5 Il vettore della malaria: *Anopheles gambiae*

Il genere *Anopheles* conta più di 400 specie diverse ma di queste solo 30 risultano essere importanti come vettori di malaria. La specie *A. gambiae* è la più efficiente in quanto presenta una maggiore suscettibilità al parassita *Plasmodium* e manifesta la sua preferenza per l'ospite uomo rispetto ad altri animali (antropofilo). Solo le zanzare adulte femmine si nutrono di sangue ed effettuano un pasto ogni 3-4 giorni, grazie al quale ottengono nutrienti per la maturazione delle uova. Le femmine, inoltre, presentano un tempo di vita medio di circa una o due settimane, rispetto ai maschi che sopravvivono solo pochi giorni. In genere, risultano essere maggiormente attive al crepuscolo e durante la notte, anche se negli ultimi anni alcune popolazioni di zanzare hanno sviluppato la capacità di modificare il ritmo giorno-notte [7].

1.5.1 Ciclo di vita e morfologia delle zanzare *Anopheles gambiae*

La vita delle zanzare *A. gambiae* è suddivisa in quattro fasi principali: uovo, larva, pupa e adulto.

Le uova sono deposte singolarmente dalla femmina uno o due giorni dopo il pasto direttamente sulla superficie dell'acqua. Presentano una lunghezza di circa 0,48 mm e non sono resistenti alla siccità. Dopo circa due o tre giorni dalla deposizione, le uova si schiudono e danno vita alle larve. In totale vi sono quattro stadi larvali, durante i quali aumentano le dimensioni della larva fino a raggiungere una dimensione di 5-6 mm nell'ultimo stadio. Esse si nutrono di alghe,

microorganismi e particelle organiche che si trovano nell'acqua. Successivamente le larve si trasformano in pupe che presentano una forma del corpo singolare, estremamente curvata, in cui capo e torace sono fusi. Le pupe sono molto mobili, non si nutrono e preferiscono spostarsi rapidamente lungo la superficie del loro habitat acquatico. Dopo un massimo di 5 giorni, la pupa fuoriesce dall'ambiente acquatico e si trasforma in adulto.

L'intero ciclo di vita dura 10-11 giorni ma la temperatura può influenzarne la durata prolungandola a circa tre settimane in condizioni fredde.

La specie *Anopheles* adulta presenta una postura tipica che la distingue da tutte le altre. A differenza di altre specie che mantengono il corpo parallelo alla superficie, le *Anopheles* presentano una posizione di riposo singolare con l'addome sollevato in aria e un'inclinazione in avanti di 30-45 gradi, con il capo più vicino alla superficie stessa. Le zanzare *A. gambiae* sono di piccole o medie dimensioni. Il corpo ha un colore che generalmente varia dal grigio al marrone chiaro, mentre le ali presentano un aspetto squamoso con macchie chiare di colore giallo o bianco e una lunghezza che mediamente varia tra 2,8 e 4,4 mm (Figura 4) [7] [11].



Figura 4. *Anopheles gambiae* durante un pasto [7].

Capitolo 2

Prevenzione, strategie di contenimento e trattamento della malaria

La prevenzione della malaria comporta l'applicazione di misure di controllo del vettore, l'assunzione di farmaci preventivi, nota come chemioterapia preventiva, e la somministrazione di vaccini. Gli obiettivi sono la riduzione del numero di casi e decessi causati da un'elevata trasmissione e l'eliminazione della malattia a livello locale e/o regionale.

2.1 Le strategie di controllo del vettore

Le strategie di controllo possono essere distinte in azioni che impediscono o limitano il contatto tra l'insetto e l'ospite e azioni che riducono la densità delle zanzare.

Gli interventi che riducono l'interazione prevedono:

- l'utilizzo di abiti chiari e lunghi che coprono il corpo in particolare nelle ore notturne;
- l'utilizzo di repellenti sulla cute esposta o sugli indumenti;
- l'uso di zanzariere su porte e finestre;
- l'utilizzo di zanzariere da letto impregnate con insetticida (ITN);
- il soggiorno all'aperto preferibilmente nelle ore in cui le zanzare *Anopheles* non sono attive;

Gli interventi che mirano a ridurre la densità delle zanzare, invece, prevedono:

- l'utilizzo di insetticidi ad azione residua nelle abitazioni (IRS);
- la riduzione dei focolai larvali attraverso bonifiche ambientali;
- l'uccisione delle larve nelle acque stagnanti attraverso l'uso di prodotti larvicidi;
- l'introduzione di controlli biologici come predatori naturali;

La profilassi comportamentale garantisce un certo livello di protezione ed è risultata essere efficace anche in Africa, dove la malaria è endemica, riducendone l'impatto sulla popolazione residente [12].

2.1.1 Le zanzariere da letto impregnate con insetticidi

Le zanzariere da letto impregnate con insetticidi (ITN) fungono da barriera protettiva per le persone che dormono al di sotto. Hanno forme, colori e dimensioni diverse e i principali materiali utilizzati per la realizzazione sono il poliestere, il polietilene e il polipropilene.

L'Organizzazione Mondiale della Sanità distingue due tipi di ITN in base all'insetticida che viene utilizzato: quelle capaci di uccidere le zanzare e quelle in grado di sterilizzarle o ridurre la fecondità. Alla prima categoria appartengono le zanzariere che combinano il piretroide con il pirrolo e quelle a base di piretroidi e/o piperonil butossido (PBO), in grado di inibire gli enzimi metabolici delle zanzare. Alla seconda categoria, invece, appartengono le zanzariere trattate con piretroide e piriproxifene, un insetticida regolatore della crescita degli insetti capace di interferire sulla riproduzione e sullo sviluppo delle zanzare femmine. L'utilizzo delle ITN ha determinato una riduzione del carico della malaria e potrebbe garantire una protezione per tutta la comunità, se solo più della metà dei membri ne facessero uso. Tuttavia, l'uso improprio di questi strumenti di protezione potrebbe comportare un rischio per la salute umana. Alcuni effetti dannosi potrebbero essere: danni psicomotori, malformazione dei feti e cancro. In particolare, studi sulla permetrina hanno dimostrato come vi sia una correlazione tra la cattiva gestione di questo insetticida e l'insorgenza di effetti avversi sul sistema nervoso e sulla pelle, manifestando prurito e bruciore, soprattutto in bambini e neonati. La potenziale tossicità è associata anche all'uso prolungato delle ITN e all'avvelenamento indiretto quando utilizzate come reti da pesca. L'immersione in acqua, infatti, comporta l'emissione di sostanze dannose per l'ambiente acquatico e l'accumulo di piretroidi nei pesci destinati al consumo umano. [13] [14] [15]

2.1.2 Gli insetticidi ad azione residua nelle abitazioni.

L'irrorazione residua interna (IRS) delle abitazioni prevede l'applicazione di insetticidi sulle pareti o su altre superfici di appoggio delle zanzare, quali ad esempio grondaie o soffitti. È possibile distinguere tra insetticidi da contatto residui, la cui tossicità permane anche dopo diversi mesi, e non residui, in cui le sostanze chimiche organiche sono rapidamente degradate nell'ambiente. Il periodo minimo di efficacia, definito dall'Organizzazione Mondiale della Sanità, è di 2-6 mesi. Diversi fattori possono influire sull'efficienza e l'efficacia dell'irrorazione tra cui: il livello di formazione degli operatori, la natura dell'insetticida, i fattori ambientali, l'endemicità della malattia, i comportamenti di alimentazione e riposo dei vettori, i costi e la trascurata e/o mancata protezione dell'ambiente domestico dopo l'operazione di irrorazione. L'IRS è raccomandata prima dell'inizio della stagione in cui si verifica il maggior numero di trasmissioni di malaria e non sono consentiti più di due cicli in un anno. Nonostante l'irrorazione sia più costosa delle reti trattate con insetticidi di lunga durata (LLIN), risulta essere un intervento vantaggioso dal punto di vista economico anche nei Paesi a basso reddito e si considera la possibilità di poterle utilizzare in combinazione.

La mancata consapevolezza a livello locale da parte della popolazione e, di conseguenza, anche l'accettazione delle strategie di contenimento rappresentano ancora oggi un aspetto critico [16] [17] [18].

2.1.3 Il controllo larvale

Il controllo larvale si basa sull'utilizzo di larvicidi e sulla riduzione della fonte che prevede la rimozione o la distruzione dei siti di riproduzione delle zanzare, al fine di modificarli e renderli inadatti allo sviluppo delle larve.

La riduzione della fonte comprende sia azioni di modifica che di manipolazione ambientale. Le azioni di modifica determinano l'alterazione o l'eliminazione dei luoghi di riproduzione in modo permanente, tramite la schermatura dei contenitori d'acqua, il prosciugamento delle paludi e degli stagni o il riempimento di fossati e pozze che raccolgono l'acqua. Può essere considerato un'azione di modifica l'utilizzo della specie arborea Eucalipto, poiché grazie alla capacità di crescere velocemente permette di essiccare terreni eccessivamente paludosi. Le azioni di manipolazione ambientale, invece, riguardano un insieme di misure da ripetere nel tempo quali ad esempio lo scarico di corsi d'acqua, la pulizia dei contenitori d'acqua e la riparazione degli scarichi [19] [20].

I larvicidi vengono utilizzati direttamente sugli habitat larvali. Alcuni esempi sono le perline di polistirolo, gli oli vegetali e il farmaco ivermectina. Le perline di polistirolo e gli oli vengono applicati sulla superficie dell'acqua in maniera tale da creare uno strato galleggiante e impedire alle larve il raggiungimento della superficie per respirare, determinandone la morte.

Gli oli vegetali sono ottenuti dall'essiccazione all'ombra, per una settimana, di parti di alcune piante prima dell'idro-distillazione. In particolare, gli oli ottenuti dalle foglie delle piante *Hyptis suaveolens*, *Ocimum gratissimum*, *Nicotiana tabacum*, *Ageratum conyzoides* e la buccia del frutto di *Citrus sinensis* hanno dimostrato un'elevata capacità tossica contro gli stadi larvali e le femmine adulte di *Anopheles gambiae*. Gli oli, inoltre, sono biodegradabili, facilmente accessibili e convenienti in quanto meno dannosi per l'ambiente [21].

A seguito della ridotta sopravvivenza delle zanzare adulte dopo un pasto di sangue prelevato da un ospite trattato con ivermectina è stato condotto uno studio sul farmaco. Si tratta di un antiparassitario utilizzato sia in medicina veterinaria che umana e può contaminare l'habitat larvale anche attraverso gli escrementi del bestiame trattato. Lo studio ha dimostrato che l'esposizione all'ivermectina di *A. gambiae* ha determinato una riduzione della sopravvivenza larvale dose-dipendente. Lo stadio in cui le larve erano esposte non è risultato essere influente

sui risultati e non sono stati registrati effetti sul numero di uova deposte e sul tasso di schiusa delle stesse dopo l'esposizione [22].

2.2 La chemioterapia preventiva

La chemioterapia preventiva consiste nell'assunzione di farmaci antimalarici, da soli o in combinazione, da parte delle popolazioni a rischio per prevenire l'infezione e le conseguenze ad essa associate. La terapia preventiva comprende:

- Chemioprevenzione stagionale della malaria (SMC): include la somministrazione in combinazione dei farmaci sulfodossina/pirimetamina e amiodachina durante la stagione di massima diffusione malarica. È attualmente utilizzata in alcune aree dell'Africa occidentale, in cui la trasmissione della malaria è limitata. Non è raccomandata in zone dell'Africa in cui vi è un'elevata trasmissione della malattia tutto l'anno [23].
- Trattamento preventivo intermittente per le donne in gravidanza (IPTp): prevede la somministrazione del farmaco antimalarico sulfodossina/pirimetamina (SP) durante le visite di assistenza prenatale. Lo scopo della somministrazione di ogni dose è la soppressione di un'eventuale infezione asintomatica, che potrebbe contribuire all'insorgenza dell'anemia materna e ad una più elevata probabilità che il neonato abbia un basso peso alla nascita. È raccomandato un minimo di tre dosi (IPTp3+) e una copertura pari almeno all'80% delle donne in gravidanza [24] [25].
- Trattamento preventivo intermittente per neonati (IPTi): prevede la somministrazione del farmaco antimalarico sulfodossina/pirimetamina (SP) al momento della vaccinazione di routine ai neonati a rischio di contrarre l'infezione. Il trattamento preventivo è esteso anche ai bambini di età inferiore ai 5 anni (IPTe) e ai bambini in età scolare (IPTsc) compresi tra i 5 e i 15 anni ed è erogato attraverso le scuole. L'IPT nei bambini è somministrata, in genere, mensilmente soprattutto durante la stagione di massima diffusione malarica [24] [25] [26].
- Somministrazione di farmaci di massa (MDA): consiste nella somministrazione di farmaci a tutta la popolazione. È indicato nelle zone a bassa, media ed elevata trasmissione malarica. Vi è, tuttavia, la necessità di indagare eventuali complicazioni a lungo termine della MDA [26].
- Chemioprolifassi: è raccomandata in caso di spostamento e soggiorno da zone non endemiche a zone endemiche. I farmaci non garantiscono una protezione al 100% dalla malattia, ma riducono in rischio di contrarla in forma grave, trattandosi di soggetti non coperti da immunità innata o acquisita. La scelta del trattamento da somministrare dipende

dal Paese di soggiorno, dal tempo di permanenza, dalla stagione e dalla salute del soggetto, specie se si tratta di bambini o di donne in gravidanza. [27]

I diversi trattamenti hanno dimostrato di ridurre significativamente la mortalità ma si raccomanda di utilizzarli in combinazioni con le strategie di contenimento per una maggiore efficacia.

2.3 I vaccini

Il maggior numero di infezioni malariche è causato dalla specie *Plasmodium falciparum*, il cui complesso ciclo di riproduzione ha portato, nel corso degli anni, all'esplorazione di vari approcci vaccinali mirati alle diverse fasi del ciclo di vita del parassita. Gli studi sullo sviluppo di un vaccino contro la malaria sono in corso dagli anni '60 e hanno portato alla definizione di tre grandi categorie [28] [29]:

- Vaccini pre-eritrocitari (PEV): mirati ai parassiti allo stadio sporozoitico e allo stadio epatico e hanno lo scopo di prevenire l'infezione;
- Vaccini eritrocitari (BSV): che limitano la crescita di parassiti allo stadio sanguigno;
- Vaccini bloccanti la trasmissione (TBV): che prendono di mira i parassiti allo stadio sessuale o le zanzare; dunque, riducono/bloccano la trasmissione ed eliminano la malattia attraverso l'immunità di gregge (Figura 5);

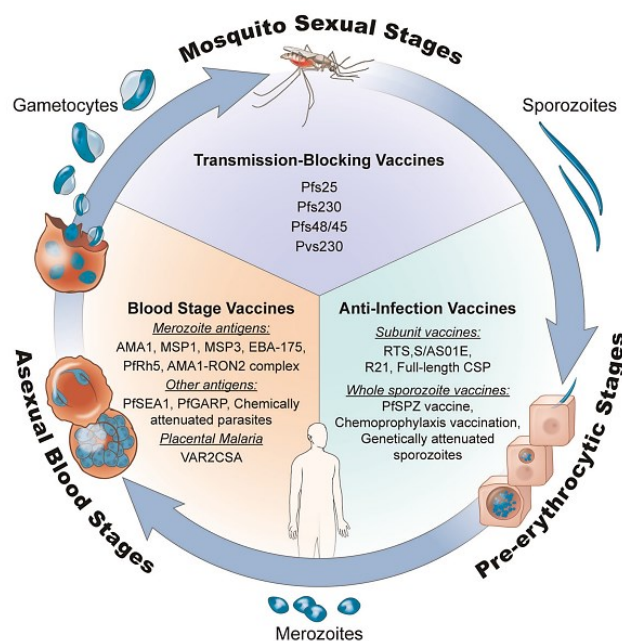


Figura 5: Categorie di vaccini mirati alle diverse fasi di ciclo di vita del parassita [29].

2.3.1 I vaccini pre-eritrocitari

I vaccini pre-eritrocitari (PEV) prendono di mira gli sporozoiti e i parassiti allo stadio epatico e sono nati da studi condotti decenni fa in cui si osservò che sporozoiti attenuati dalle radiazioni (RAS) erano in grado di conferire immunità sia agli animali che agli uomini. Gli obiettivi dei PEV sono: lo sviluppo di anticorpi contro gli antigeni degli sporozoiti, così da eliminarli dal flusso sanguigno e bloccare l'invasione delle cellule epatiche, o l'intervento delle cellule T del sistema immunitario, pronti ad attaccare gli epatociti infetti. Sono definiti anche vaccini anti-infezione (AIV).

Tra i vaccini pre-eritrocitari distinguiamo quelli che utilizzano sporozoiti interi (PfSPZ) e quelli che fanno uso della proteina circumsporozoite (CSP).

Il termine PfSPZ fa riferimento agli sporozoiti (SPZ) della specie *P. falciparum* (Pf) che vengono prelevati dalle ghiandole salivari delle zanzare e sono attenuati in diversi modi. A seconda dell'approccio utilizzato per l'attenuazione conferiscono il nome al vaccino candidato:

- Vaccino PfSPZ: sporozoiti attenuati dalle radiazioni. È uno dei vaccini candidati più promettenti e mostra un'incoraggiante protezione se somministrato per via endovenosa, risultando sicuro ed efficace sia nei neonati che nei bambini;
- Vaccino PfSPZ-CVac: chemioattenuazione ottenuta mediante la somministrazione con farmaci antimalarici;
- PfSPZ-GA1: attenuazione genetica mediante delezioni di geni necessari allo sviluppo allo stadio epatico [28] [29] [30].

La proteina CSP è una proteina secreta dagli sporozoiti della specie *P. falciparum*, rappresenta il principale bersaglio antigenico ed è inclusa nel vaccino antimalarico oggi più avanzato, il vaccino RTS,S/AS01 e nel vaccino R21/MM [28] [29].

2.3.1.1 Il vaccino RTS,S/AS01

Il vaccino RTS,S/AS01 include una forma troncata della proteina CSP formata dalla regione C-terminale (contenente epitopi di cellule T, da qui "T") e dalla regione centrale ripetuta (da qui "R"), che comprende una sequenza amminoacidica ripetuta di asparagina-alanina-asparagina-prolina (NANP). Il frammento antigenico è poi fuso con l'antigene di superficie dell'epatite B (HbsAg, da qui "S"), mentre il termine AS01 fa riferimento all'adiuvante, a base di liposomi, utilizzato per potenziare la risposta del sistema immunitario [29] [31].

RTS,S/AS01, commercializzato come Mosquirix, è il primo e l'unico vaccino raccomandato e prequalificato dall'OMS dal 6 Ottobre 2021. Lo studio più importante sul vaccino è stato

condotto dal 2009 al 2011, i cui risultati promettenti hanno portato all'approvazione favorevole da parte dell'Agenzia europea per i medicinali (EMA) nel 2015 e l'avvio, nel 2019, di studi su larga scala in Ghana, Kenya e Malawi. Il vaccino attualmente è stato approvato solo per i bambini di età inferiore ai 2 anni che vivono in regioni in cui la trasmissione della malattia è medio/alta. La somministrazione della prima dose avviene intorno ai 5-6 mesi e il richiamo all'età di 2 anni. Ha dimostrato di ridurre l'incidenza della malaria, i casi gravi, i ricoveri e la necessità di trasfusione, conferendo una maggiore protezione se la somministrazione è accompagnata dall'utilizzo di altre strategie di controllo. Nonostante ciò, vari studi hanno riportato che RTS,S/AS01 presenta importanti limitazioni e l'efficacia risulta essere ancora modesta e di breve durata. Nei bambini di età compresa tra i 5 e i 17 mesi l'efficacia è stata del 55% dopo tre dosi, a distanza di 12 mesi, e vicina allo zero entro i 18 mesi. La dose di richiamo dopo i 18 mesi ha conferito un'immunità con un'efficacia del 36% contro la malaria sintomatica e del 29% contro la malaria grave. Nel caso dei neonati di età compresa tra i 6 e i 12 mesi, l'efficacia è stata considerata troppo bassa [31]. Inoltre, sono stati segnalati importanti reazioni avverse tra cui meningiti, malaria cerebrale, polmonite, anemia e convulsioni febbrili in tutte le fasce d'età compresi i neonati [30]. L'obiettivo è migliorarne l'efficacia e la longevità tramite il superamento di importanti limitazioni e lo sviluppo di nuove strategie [31] quali :

- l'utilizzo di ulteriori alleli CSP, poiché è presente un solo allele con numerosi polimorfismi principalmente nella regione C-terminale;
- l'inclusione di ulteriori antigeni sporozoitici o epitopi CSP;
- la combinazione di RTS,S con antigeni merozoitici dello stadio sanguigno per aumentare la protezione contro la malattia;
- l'inclusione di un antigene che generi anticorpi che bloccano la trasmissione.

Sono stati segnalati molti fattori che influiscono su una completa copertura vaccinale come ad esempio: la negazione, da parte della popolazione, di campagne di vaccinazioni dovuta ad una mancata conoscenza sui vaccini, la loro efficacia e il loro meccanismo d'azione; la fornitura inadeguata di servizi di immunizzazione; la distanza dalle strutture di vaccinazione e le strutture sanitarie inadeguate. Inoltre, anche la cultura tradizionale, secondo cui le decisioni sanitarie spettano ad un singolo membro della famiglia, si è rivelata essere un ostacolo [30].

2.3.1.2 Il vaccino R21/MM

Il vaccino R21/MM è un vaccino pre-eritrocitario che ha dimostrato, negli studi finora condotti, una maggiore immunogenicità rispetto al RTS,S/AS01. La particella R21 contiene solo la proteina di fusione CSP-HbsAg senza la necessità di un eccesso di antigene dell'epatite B

richiesto, invece, per RTS,S. L'adiuvante utilizzato, in questo caso, è Matrix-M (MM). I dati hanno dimostrato nei bambini di età inferiore ai 2 anni, precedentemente esposti alla malaria, un'efficacia di almeno il 75% dopo la somministrazione di 3 dosi. Tale efficacia è aumentata a seguito di una dose di richiamo somministrata 12 mesi dopo. Non sono state segnalate importanti reazioni avverse ed è stato autorizzato per l'uso in tre Paesi: Ghana, Nigeria e Burkina Faso. Sono, tuttavia, necessari studi clinici controllati randomizzati più ampi per valutarne l'efficacia in una popolazione più vasta (Figura 6) [32] [33].

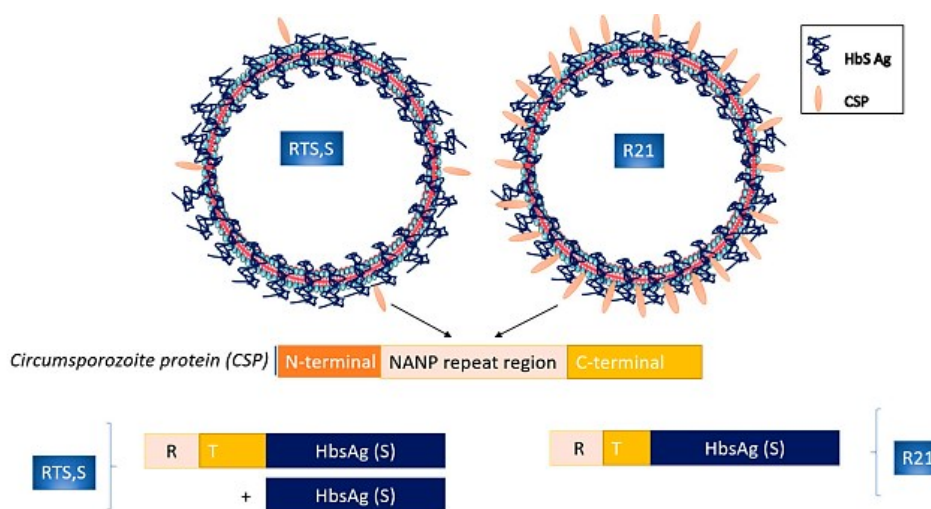


Figura 6: Rappresentazione e confronto delle strutture dei vaccini RTS,SA01 e R21/MM [32].

2.3.2 I vaccini eritrocitari

I vaccini eritrocitari o vaccini allo stadio sanguigno (BSV) si basano su un meccanismo d'azione che tenta di bersagliare i parassiti asessuali (merozoiti) così che possa essere bloccata l'invasione dei globuli rossi. In particolare, vengono presi di mira gli antigeni altamente espressi sulla superficie di merozoiti. Dagli studi clinici sono emersi risultati sconcertanti e i principali limiti sono associati: al passaggio di merozoiti tra gli eritrociti, che avviene in un tempo breve e fornisce un accesso limitato agli anticorpi neutralizzanti, al polimorfismo antigenico e a percorsi di invasione ridondanti [28] [29] [32].

Un tipo particolare di vaccini BSV entrato nella pratica clinica è rappresentato dai vaccini contro la malaria placentare (PMV). I PMV prendono di mira i parassiti che legano la condroitin solfato A (CSA), il principale recettore presente negli spazi sanguigni intervillosi della placenta. Si tratta di approccio mirato, grazie al quale le donne sono in grado di acquisire anticorpi

naturalizzanti nelle gravidanze seguenti quando diventano resistenti alla malaria placentare [29].

2.3.3 I vaccini bloccanti la trasmissione

I vaccini bloccanti la trasmissione (TBV) mirano ai parassiti allo stadio di zanzara, puntando agli antigeni di superficie dei gameti e degli zigoti. Gli antigeni candidati sono: *Pfs230* che forma un complesso con *Pfs48/45* e svolge un ruolo nella fertilità del gamete maschile, *Pfs25* e *Pfs28* che sono le principali proteine di superficie degli zigoti espresse solo dopo la fecondazione nella zanzara ospite e *Pfs47* che è l'antigene di superficie del gamete femminile. I vaccini TBV sono ancora nella fase 1 della sperimentazione clinica, in particolare i candidati *Pfs25* hanno portato alla generazione di risposte immunitarie deboli e di durata limitata. L'obiettivo è l'induzione di livelli di anticorpi efficienti mantenuti per almeno una stagione di trasmissione [28] [29] [34].

2.4 Lo sviluppo di una popolazione resistente

Le zanzariere impregnate con insetticida (ITN), comprese le reti trattate con insetticida a lunga durata (LLIN), e l'irrorazione residua nelle abitazioni (IRS) sono le principali strategie di controllo utilizzate per ridurre il numero di casi e decessi di malaria. Gli interventi messi in atto, però, mostrano un plateau nell'efficienza e non sono sufficienti ad arrestare la trasmissione della malattia, tanto da indurre l'Organizzazione Mondiale della Sanità a parlare di "trasmissione residua". La trasmissione residua è indicata come "*la persistenza della trasmissione della malaria a seguito dell'implementazione nel tempo e nello spazio di un programma contro la malaria ampiamente efficace*" [35]. La trasmissione residua può essere influenzata da diversi fattori, alcuni dei quali sono: mancata accettazione e utilizzo errato degli interventi, resistenza ai farmaci, resistenza agli insetticidi e adattamento del vettore alle strategie di controllo [36] [37].

2.4.1 Resistenza ai farmaci

La resistenza ai farmaci antimalarici rappresenta una delle principali sfide degli ultimi anni ed è stata confermata in due specie: *P. falciparum* e *P. vivax*. Il *P. falciparum* è risultato essere resistente ai farmaci sulfadossina/pirimetamina (SP), artemisinina, meflochina, alofantrina e chinino. La resistenza è la conseguenza di un uso eccessivo e improprio di farmaci che ne compromette l'efficacia. In particolare, la resistenza alla SP, somministrata come trattamento

preventivo intermittente (IPT), soprattutto nelle donne incinte (IPTp) e nei neonati (IPTi), è stata associata a un polimorfismo a singolo nucleotide (SNP) in due geni diversi ed è diffusa in tutto il mondo. Nel Sud-Est asiatico e in alcune regioni dell’Africa, è stata confermata anche la resistenza all’artemisinina, il principale farmaco della terapia ACT (terapia di combinazione con l’artemisinina). Essa si basa su un derivato dell’artemisinina e un secondo farmaco antimalarico ed è raccomandata dall’OMS come trattamento di prima scelta contro la malaria. Attualmente i tentativi di contrastare la farmaco resistenza si basano sull’identificazione di nuovi farmaci e sulla modifica chimica o la combinazione dei farmaci già disponibili in commercio, nel tentativo di prevenire o ritardare la resistenza [24] [37] [38] [39].

2.4.2 Resistenza agli insetticidi

La resistenza agli insetticidi è definita come la capacità dell’*Anopheles* di sopravvivere ad una dose elevata di insetticida e si verifica quando il suo utilizzo avviene per un lungo periodo di tempo. I diversi meccanismi molecolari alla base della resistenza sono: l’incapacità dell’insetticida di legare i recettori target della zanzara, a causa dell’insensibilità del sito bersaglio, e l’ispessimento o la modifica della cuticola che ne determina una minore penetrazione. I principali insetticidi utilizzati per l’irrorazione residua delle abitazioni sono a base di piretroidi, organoclorurati diclorodifeniltricloroetano (DDT), organofosfati e carbammati, mentre quelli utilizzati per le reti a lunga durata sono a base di piretroidi. Un aumento della resistenza verso più insetticidi è stato osservato quando questi condividono lo stesso sito target. L’*A. gambiae* è risultata resistente ai piretroidi nell’Africa occidentale, alla permetrina e DDT in Burkina Faso e al DDT e ai piretroidi nell’Africa sub-sahariana. Il problema associato alla resistenza agli insetticidi è determinato non solo dalla sopravvivenza ma anche alla modifica dei comportamenti dei vettori malarici e alla comparsa di nuove popolazioni di zanzare. Le strategie di gestione per combattere la resistenza si basano sulla combinazione di insetticidi maggiormente efficaci [24] [40].

Capitolo 3

Il ruolo dell'ingegneria genetica nel controllo della malaria

Le strategie di controllo dei vettori utilizzate per l'eliminazione della malaria a livello locale non sono sufficienti a determinarne l'eradicazione globale. La resistenza a farmaci e insetticidi e l'adattamento del vettore non sono gli unici aspetti che rendono complesso il controllo delle zanzare. Molte considerazioni devono essere fatte anche dal punto di vista sociale ed economico. Si pensa, infatti, che la povertà, le scarse condizioni igieniche, i conflitti civili nei Paesi endemici, l'inaccessibilità da parte della popolazione a farmaci nuovi e l'incapacità delle infrastrutture esistenti di contrastare la portata del problema, possano addirittura peggiorare la situazione e determinare un aumento dei costi degli attuali interventi di prima linea. Nuove strategie per il controllo delle zanzare e l'eradicazione delle malattie trasmesse da vettori (VBD) si basano sul controllo genico. Si tratta di strategie alternative, efficaci, più economiche e sostenibili [41] [42].

3.1 Il controllo genico

Il concetto di controllo genico risale agli anni '30 -'40 e prevede il rilascio di individui con tratti genetici desiderabili e deleteri, così da diffondere questo tratto nella popolazione selvatica tramite l'accoppiamento e l'ereditarietà. I tratti genetici inseriti possono portare alla sostituzione o alla soppressione della popolazione. La sostituzione (o modificazione) prevede la riduzione della capacità del vettore di diffondere il parassita portando ad una diminuzione dell'incidenza della malattia ma non della popolazione delle zanzare. La soppressione, invece, comporta la compromissione della capacità del vettore di riprodursi, di conseguenza la popolazione sarà così piccola da non poter sostenere la trasmissione della malattia. Un'ultima osservazione va fatta sulla persistenza della modifica nella popolazione selvatica distinguendo, dunque, tra strategie autolimitanti, in cui la modifica tende a scomparire, e strategie autosufficienti, grazie alle quali, invece, la modifica permane e si diffonde nella popolazione. Le strategie di controllo geniche finora maggiormente utilizzate e definite autolimitanti sono la tecnica dell'insetto sterile (SIT) e il rilascio di insetti portatori di un gene letale dominante (RIDL) [41] [43].

3.1.1 Tecnica dell'insetto sterile (SIT) e rilascio di insetti portatori di un gene letale dominante (RIDL)

La SIT rappresenta un tentativo di modifica della biologia del vettore per il controllo delle malattie negli esseri umani. Il rilascio di insetti maschi sterilizzati tramite raggi X porta ad un declino della popolazione bersaglio. La sterilizzazione è ottenuta mediante irradiazione o metodi di sterilizzazione a base chimica, che inducono mutazioni nel DNA delle cellule germinali. Il successo di questa strategia è stato osservato solo in alcuni insetti quali ad esempio la mosca mediterranea della frutta (*Ceratitidis capitata*) e le zanzare *Aedes aegypti*, responsabili della Dengue e della febbre gialla. In altre specie di zanzare la tecnica ha mostrato risultati deludenti. I principali limiti di questo approccio riguardano:

- la necessità di un allevamento, su scala industriale, di un numero di maschi sterili maggiori rispetto alla popolazione selvatica locale;
- investimenti sostenuti per il rilascio costante del vettore;
- l'alta competitività richiesta dai maschi sterili contro i maschi selvatici;
- la compromissione della competitività generale dell'accoppiamento maschile a causa dei metodi di sterilizzazione;

Nel tentativo di superare alcune di queste limitazioni sono stati compiuti dei progressi sviluppando metodi di sterilizzazione più precisi che preservano la fertilità maschile ma uccidono la prole negli stadi di sviluppo post-embrionale, approcci che permettono l'espressione di fattori che promuovono l'apoptosi nei testicoli e strategie che provocano la sterilizzazione nei maschi e l'uccisione delle femmine. Nessuno di questi metodi, tuttavia, è stato applicato nelle zanzare *Anopheles gambiae* [43] [44].

L'approccio RIDL si propone di superare le limitazioni della SIT in relazione alla mancanza di un metodo efficiente per la selezione dei maschi e alla letalità causata dalle radiazioni. Sfrutta l'allevamento in massa di ceppi RIDL e il rilascio di soli insetti maschi portatori di transgeni letali dominanti femminili, quindi sesso-specifici. Le femmine vengono rimosse meccanicamente o utilizzando un sistema letale. I maschi vengono poi rilasciati per competere con i maschi wild-type e accoppiarsi con le femmine selvatiche. L'accoppiamento del maschio portatore del transgene e la femmina selvatica genera una progenie che non è in grado di sopravvivere oltre lo stadio di pupa (Figura 7). La morte è determinata dal fatto che la prole, oltre al transgene, eredita una proteina di fusione transattivatrice che è reprimibile solo con la tetraciclina. L'assenza della tetraciclina in natura provoca la morte della prole e la diminuzione della popolazione locale di zanzare. È necessario, anche in questo caso, il rilascio periodico del vettore così da raggiungere una soglia tale da reprimere la popolazione. Sebbene l'approccio

complessivo sembra essere efficiente con le popolazioni di zanzare isolate del genere *Anopheles*, vi è l'impossibilità del gene letale di spostarsi su lunghe distanze. Questo svantaggio lo rende un approccio non adatto al controllo della malaria [42] [45] [46].

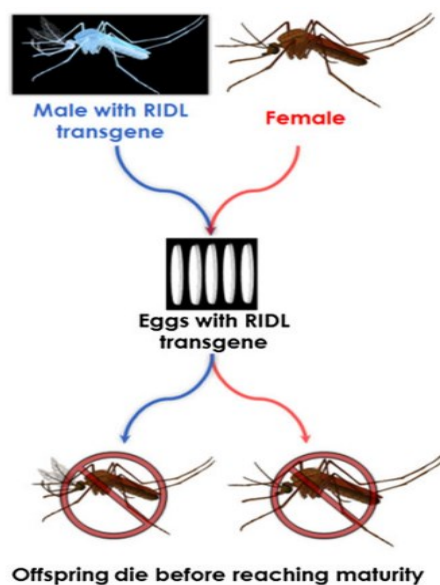


Figura 7: Meccanismo di RIDL nella popolazione di zanzare [46].

3.2 Gene drive

L'elevato numero di zanzare transgeniche necessario rispetto a quelle selvatiche e il bisogno di raggiungere una frequenza apprezzabile nell'area target impongono il ricorso ad una forma di controllo genetico innovativa definita gene drive. I gene drive sono elementi genetici che, negli organismi che si riproducono sessualmente, sono trasmessi a frequenze super-mendeliane (>50%). L'eredità mendeliana prevede che il 50% della progenie riceva una copia di ciascun allele² parentale senza alcuna pressione selettiva, per cui la sua frequenza rimane costante nella popolazione, e nel corso del tempo, tende a perdersi soprattutto se il transgene comporta un costo in termini di fitness³. Al contrario, in un sistema super-mendeliano, i geni vengono trasmessi a oltre il 50% della progenie consentendo alla modifica di essere diffusa rapidamente dopo poche generazioni (Figura 8) [41] [45] [47]. Il funzionamento del sistema super-mendeliano verrà spiegato in dettaglio nei paragrafi successivi.

L'alterazione della propria eredità, dunque, permette ai gene drive di diffondersi in maniera autonoma e in un arco di tempo molto breve all'interno di una popolazione, anche a partire da

² Un allele è una variante di un gene che controlla lo stesso tratto. Gli alleli si presentano solitamente in coppia. In base alla quantità di copie di alleli richieste per osservare il fenotipo corrispondente ad un dato allele possono essere dominanti o recessivi.

³ Fitness: indica il successo riproduttivo di un individuo.

una frequenza inizialmente molto bassa. È possibile, infatti, distinguere due categorie di gene drive:

- Gene drive “locali” (detti anche ad alta soglia o spazialmente confinati): richiedono il rilascio di molti individui al di sopra di una certa frequenza per la diffusione super-mendeliana. Se la frequenza iniziale è al di sotto della soglia, l’allele è eliminato dalla popolazione nel tempo.
- Gene drive “globali” (detti anche a bassa soglia): richiedono il rilascio di pochi individui per la diffusione super-mendeliana. Hanno una soglia di introduzione anche pari a zero e si diffondono rapidamente [48].

L’applicazione dei gene drive confinati nello spazio può essere utile se si desidera la modifica regionale e l’eliminazione da un habitat di popolazione, senza alcuna influenza sulla specie [49].

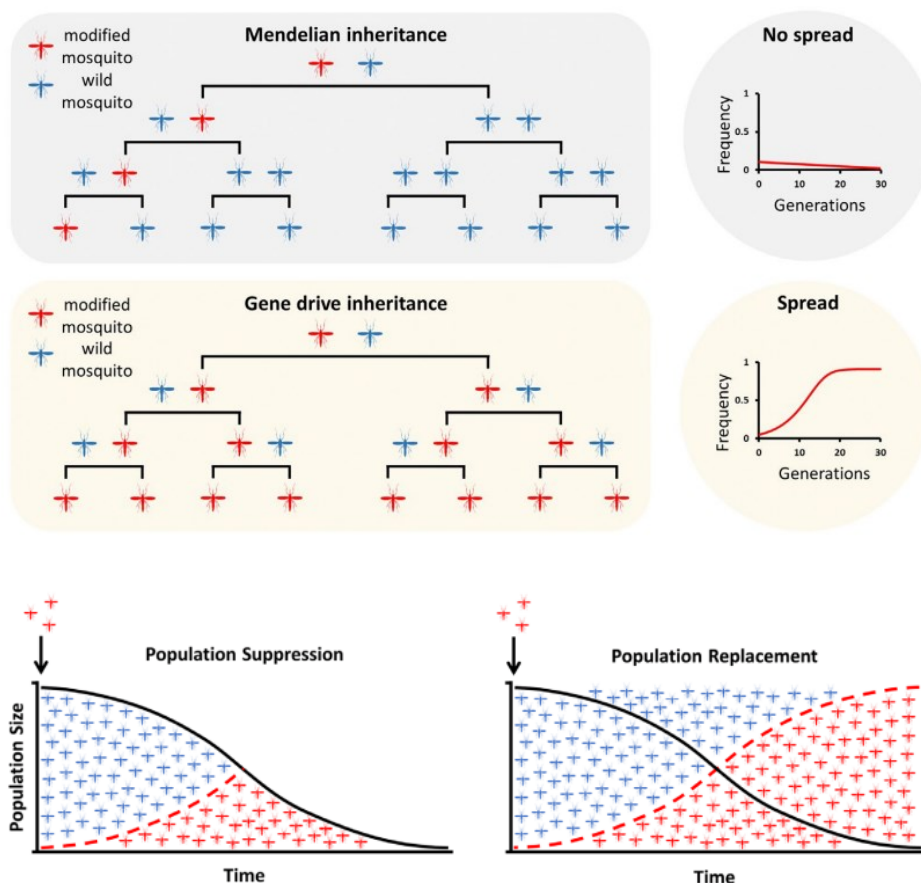


Figura 8: Differenza tra ereditarietà mendeliana ed ereditarietà super-mendeliana (in alto) e differenza tra soppressione e sostituzione della popolazione (in basso) [41].

3.2.1 Requisiti del gene drive

Un sistema gene drive necessita di numerose caratteristiche affinché possa essere efficiente nella trasmissione di tratti genetici che interferiscono con la fertilità, vitalità o capacità del vettore. Un gene drive dovrebbe:

- Garantire uno stretto legame (fisico, funzionale o entrambi) tra il meccanismo di guida e il tratto da guidare. Un legame funzionale è maggiormente desiderato in quanto risulta meno influenzato da ricombinazione, instabilità e degradazione del DNA.
- Essere robusto per contrastare un'eventuale riduzione della competitività nell'accoppiamento e fertilità;
- Poter trasportare frammenti di DNA relativamente grandi compresi il gene effetto, sequenze regolatrici e altri elementi che ne controllano le caratteristiche;
- Garantire un minimo effetto collaterale sugli organismi bersaglio e non. Questo implica la conoscenza della biologia del meccanismo effetto;
- Garantire che l'attività del gene effetto e del sistema sia limitato alla popolazione bersaglio;
- Essere applicabile ad altri vettori.

Tali caratteristiche richiedono una conoscenza della genetica, della biologia, dei comportamenti della popolazione bersaglio e delle modalità di trasmissione della malattia. In particolare, è necessario comprendere i meccanismi di accoppiamento e di funzionamento del sistema gene drive che si desidera utilizzare, poiché meccanismi differenti possono influenzare in maniera diversa l'accoppiamento tra un vettore wild-type e un vettore modificato.

3.2.2 Impatto epidemiologico di un sistema gene drive

Durante un'epidemia di una malattia infettiva, un parametro fondamentale è l'indice R_0 definito come numero di riproduzione di base. Il parametro indica il numero medio di soggetti che possono essere contagiati da un individuo infetto in una popolazione suscettibile, mai venuta a contatto con il nuovo agente patogeno. Un indice $R_0 < 1$ indica che l'epidemia può essere contenuta. Nel caso delle malattie trasmesse indirettamente, come la malaria, l'indice R_0 è calcolato come:

$$R_0 = \frac{bc*V}{r} \quad \text{con} \quad V = \frac{E [1 + cSX]}{cX}$$

dove

b= infettività⁴ della zanzara

c=infettività dell'uomo rispetto alla zanzara

V=capacità vettoriale delle zanzare

$\frac{1}{r}$ =il tempo necessario per eliminare l'infezione

E=numero medio di morsi infettivi che un soggetto riceve ogni anno

c=la probabilità che una zanzara possa infettarsi mordendo un essere umano infetto

S=numero di morsi umani per zanzara

X=frazione di zanzare infettate da parassiti

L'effetto dell'introduzione di un sistema di gene drive, sia per la soppressione che per la modifica della popolazione, risulta difficile da stimare anche a causa dell'elevato numero di parametri necessari per il calcolo di R_0 (Figura 9). Inoltre, c'è da considerare che fattori ambientali, umani e cicli stagionali di riproduzione delle zanzare influiscono sul calcolo dell'indice di riproduzione. La soppressione tende a ridurre la capacità vettoriale (V) delle zanzare, mentre la modifica tende a diminuire l'infettività (b). L'inclusione dei costi di fitness definisce una diminuzione nei livelli di soppressione e permette di definire una strategia di applicazione che vede prima il rilascio di impulsi di soppressione per ridurre il numero di zanzare, e poi quello di unità di modifica. Le unità di modifica, in questo caso, diffonderebbero e raggiungerebbero la piena introduzione più rapidamente grazie alla presenza di un numero minore di generazioni. La riduzione della popolazione ottenuta dalla soppressione, porta ad una diminuzione delle zanzare infette da parassiti, i quali, di conseguenza, svilupperebbero una minore resistenza agli effettori antimalarici espressi dalle zanzare modificate. Si stima che si potrebbe raggiungere un indice $R_0 < 1$ e un'eliminazione locale della malaria, se i sistemi di soppressione e modifica potessero mantenere livelli ridotti di zanzare infette per 2-3 anni nel caso del *Plasmodium falciparum* e per 5 anni nel caso del *P. Vivax* [48].

⁴ Infettività: capacità di un agente biologico di penetrare e moltiplicarsi in un organismo.

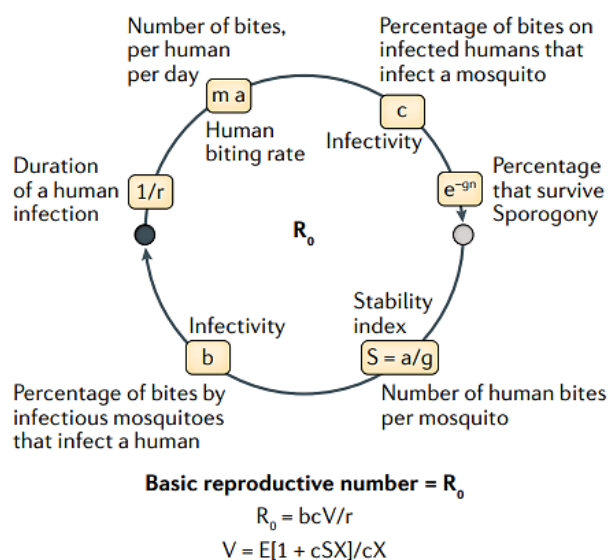


Figura 9: Numero di riproduzione di base nella malaria [48].

3.3 Sistemi gene drive

Gli elementi genetici egoistici sono abbondanti in natura e quelli maggiormente studiati sono: gli elementi trasponibili, i microrganismi ereditari, la sottodominanza genetica, l'arresto embrionale dominante con effetto materno (MEDEA), drive-meiotici, i geni dell'endonucleasi homing (HEG) e i sistemi gene drive basati su CRISPR/Cas9 (CCGD).

3.3.1 Gi elementi trasponibili

I trasposoni o elementi genetici trasponibili sono segmenti di DNA presenti nei genomi di procarioti ed eucarioti capaci di spostarsi e inserirsi in diverse posizioni del genoma. Hanno anche la capacità di colonizzare altri genomi attraverso il trasferimento orizzontale⁵. Gli elementi trasponibili vengono utilizzati per integrare costrutti genici endogeni ed esogeni modificati nei genomi dei vettori in maniera stabile ed ereditabile. Sono basati sul DNA e comprendono un gene che codifica per l'enzima trasposasi affiancato da brevi sequenze ripetute invertite di varia lunghezza che gli permettono di duplicarsi in tutto il genoma. L'elemento P è il primo elemento trasponibile studiato e responsabile di un tipo di disgenesia⁶ in *Drosophila melanogaster* [42] [50].

⁵ Trasferimento orizzontale dei geni: processo nel quale un organismo trasferisce materiale genetico ad una cellula non discendente.

⁶ Disgenesia: imperfetta formazione di un organo o di un sistema.

3.3.2 I microrganismi ereditari

Una classe di microrganismi ereditabili che possono essere utilizzati come sistemi gene drive sono i *Wolbachia*. I *Wolbachia* rappresentano una classe di batteri trasmessi dalla madre e capaci di interferire con i meccanismi di riproduzione dell'ospite. L'accoppiamento di un maschio infetto da *Wolbachia* e una femmina non infetta genera una prole non vitale. Al contrario, l'unione di un maschio (infetto o non infetto) e una femmina infetta porta ad una prole vitale e infetta [41]. Sfruttando la capacità delle femmine di accoppiarsi una sola volta, la riduzione della progenie potrebbe essere raggiunta con il rilascio di un gran numero di maschi infetti. Il meccanismo principale su cui si basa è l'incompatibilità citoplasmatica (CI). I geni *cifA* e *cifB*, presenti in *Wolbachia* sono responsabili della CI (Figura10):

- Un maschio con *cifA* e *cifB* (o solo *cifB* in alcuni organismi come, ad esempio, *A. gambiae*) che si accoppia con una femmina priva di *cifA* genera una prole non vitale (effetto "tossina");
- L'espressione di *cifA* nella femmina, invece, impedisce l'incompatibilità citoplasmatica.

La possibilità di introdurre solo questi due geni, in un sistema gene drive, anziché l'intero batterio, minimizza i cambiamenti nella popolazione e le interazioni con geni potenzialmente nuovi. Inoltre, è possibile introdurre insieme o separatamente, *cifB* come "soppressore di vitalità" e *cifA* come "salvataggio", consente un maggior controllo del sistema [49].

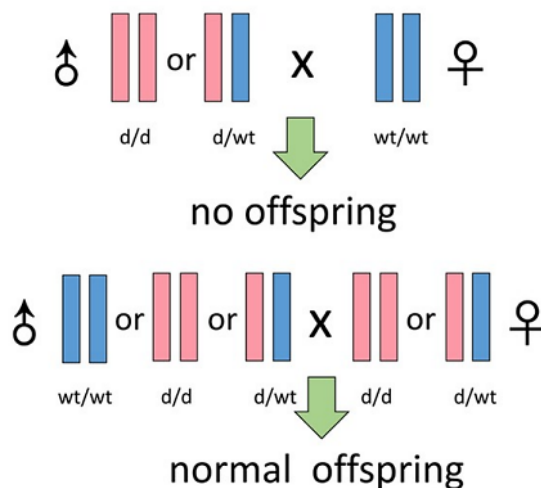


Figura 10: Meccanismo di azione di un sistema gene drive basato su *Wolbachia*, dove *d* indica l'espressione di *cifA* e *cifB* [49].

3.3.3 La sottodominanza genetica

La sottodominanza genetica è un fenomeno per cui tratti genetici in eterozigosi⁷ drive/wild-type causano infertilità o ridotta forma fisica, rispetto agli omozigoti⁸ drive o wild-type. Gli eterozigoti, quindi, hanno una minore possibilità di sopravvivere in quanto una popolazione tenderà naturalmente per l'omozigosi dell'allele che inizialmente è più frequente. Si suppone, quindi, che omozigoti wild-type (W^+/W^+) e transgenici sottodominanti (Ud/Ud) abbiano le stesse probabilità di fitness. Rilasciando omozigoti (Ud/Ud) nella popolazione wild-type, la frequenza allelica (Ud) potrebbe aumentare e sradicare gli alleli wild-type. L'eradicazione continua anche dopo l'interruzione del rilascio di alleli transgenici sottodominanti [51].

Il tratto dominante sarà in grado di invadere la popolazione solo se rilasciato al di sopra di una certa soglia richiesta. La necessità di promotori specifici, di precise sequenze target e RNAi (RNA interference) li rendono difficili da progettare [41] [49].

3.3.4 L'arresto embrionale dominante con effetto materno (MEDEA)

L'arresto embrionale dominante con effetto materno (MEDEA) indica un tipo di elementi genetici capaci di uccidere la prole non drive basandosi su un sistema di salvataggio tossina-antidoto. Si tratta di fattori autosomici che permettono la loro trasmissione alla prole, portando alla morte quella che non eredita l'allele Medea. Nelle madri portatrici di Medea la tossina e l'antidoto sono collegati. Nello specifico si verifica che:

- L'unione tra una femmina normale e un maschio eterozigote di Medea genera una prole wild-type e portatrice di Medea;
- L'unione tra una femmina eterozigote di Medea e un maschio wild-type porta alla sopravvivenza solo di prole che eredita Medea.

Questo particolare meccanismo di gene drive è stato testato in *D.melanogaster*, in cui vengono presi di mira geni materni, codificando un miRNA (microRNA), che funge da tossina, e una sequenza ricodificata insensibile al miRNA come antidoto. L'utilizzo di miRNA per la generazione di un sistema tossina-antidoto, in organismi diversi dal *D.melanogaster* risulta essere impegnativo [41] [52].

⁷ Eterozigote: organismo che presenta, per un dato gene, una coppia di alleli diversi.

⁸ Omozigote: organismo che presenta, per un dato gene, una coppia di alleli identici.

3.3.5 Il drive meiotico

Il drive meiotico è un meccanismo in cui i geni agiscono in modo tale da compromettere la generazione di gameti⁹ specifici durante il processo meiotico¹⁰. Nel caso specifico di *A. gambiae*, il cromosoma Y (determinante della mascolinità) è stato ingegnerizzato in maniera tale da trasportare un'endonucleasi¹¹ capace di tagliare e, quindi, distruggere più siti specifici sul cromosoma X, cosicché solo gli spermatozoi portatori di Y possano essere in grado di fecondare gli ovuli. Questo approccio è definito come distorsione sessuale o meccanismo Y-drive (Figura 11). Il limite principale di questa strategia è rappresentato dall'incapacità di poter determinare una espressione di geni dal cromosoma Y a causa della completa chiusura trascrizionale dei cromosomi sessuali durante la meiosi. Inoltre, un altro problema può essere associato alla presenza di varianti funzionali resistenti alle nucleasi che bloccano la diffusione del drive. Le varianti possono comprendere inserzioni¹² o delezioni¹³ generati direttamente dall'attività delle nucleasi nel sito target. Il meccanismo di distorsione sessuale funziona anche utilizzando nucleasi basate su CRISPR [41] [53].

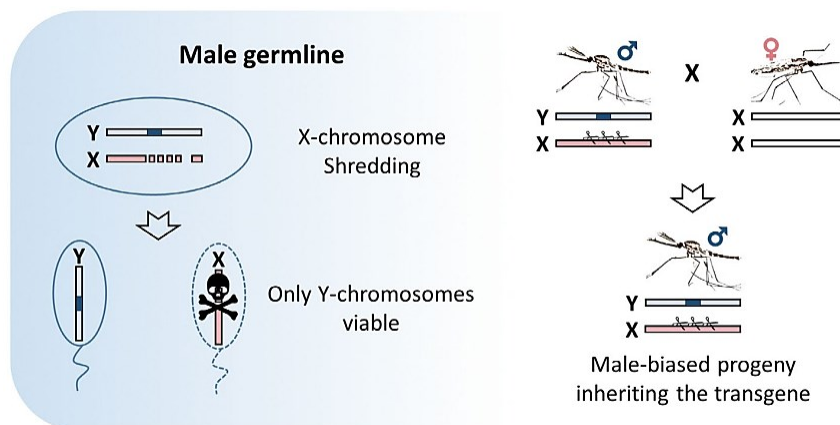


Figura 11: Meccanismo di Y-drive [41].

⁹ Gameti: cellule sessuali maschili e femminili che si uniscono durante la fecondazione per dar vita ad un nuovo organismo.

¹⁰ Meiosi: processo di divisione cellulare dei gameti.

¹¹ Endonucleasi: classe di enzimi che tagliano i legami interni in una molecola di DNA.

¹² Inserzione: aggiunta di uno o più nucleotidi.

¹³ Delezioni: perdita di uno o più nucleotidi.

3.3.6 I geni dell'endonucleasi homing (HEG)

I gene drive basati su endonucleasi sito-specifiche sono risultati essere i più efficienti nel caso delle zanzare portatrici della malaria. I geni dell'endonucleasi homing (HEG) codificano per un'endonucleasi capace di riconoscere una specifica sequenza di 15-30 bp che, di norma, si verifica una sola volta nel genoma. Dopo aver riconosciuto la sequenza bersaglio di DNA all'interno della regione di interesse, l'endonucleasi taglia la sequenza target generando nel cromosoma una rottura in entrambi i filamenti di DNA (DSB). Il taglio determina l'attivazione di un meccanismo di riparazione del DNA scelto tra meccanismo di riparazione diretta dall'omologia (HDR – *Homology Directed Repair*) o meccanismo di giunzione non omologa delle estremità (NHEJ – *Non-Homologous End Joining*). Normalmente, in un organismo diploide¹⁴, un cromosoma contenente un HEG all'interno della sequenza di riconoscimento, induce un DSB nel cromosoma omologo che, al contrario, è privo dell'HEG e contiene solo la sequenza di riconoscimento.

Tramite HDR il filamento rotto viene riparato usando come modello per la sintesi del nuovo DNA l'allele contenente l'HEG. Questo processo che permette di copiare l'HEG da un cromosoma all'altro è definito “*homing*” [47]. L'NHEJ determina l'introduzione di nucleotidi che riparano la rottura ma modificano la sequenza originale di DNA impedendo alle nucleasi di poterla riconoscere (Figura 12) [45].

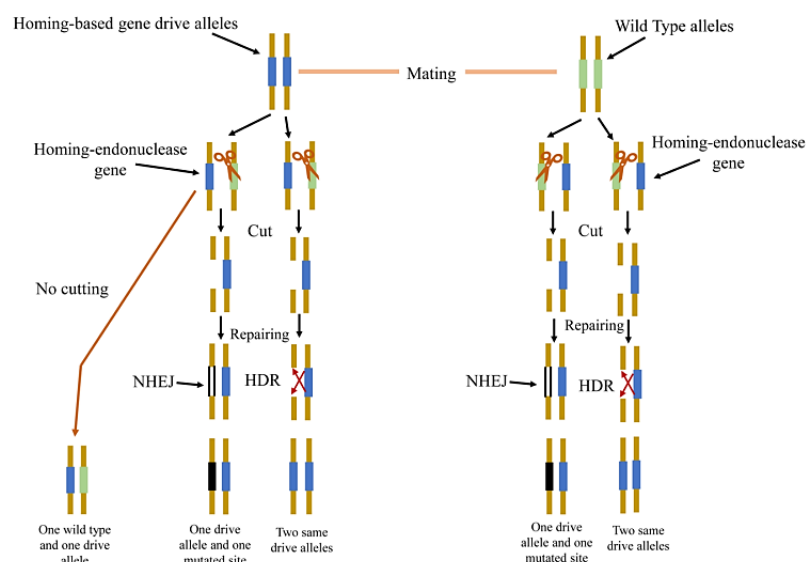


Figura 12: Gene drive basato su HEG e meccanismi di riparazione HD e NHEJ [51].

¹⁴ Organismo diploide: organismo in cui nelle cellule somatiche sono presenti due copie per ogni cromosoma.

Il meccanismo di homing è unidirezionale, in quanto una volta modificata per contenere l'HEG, la sequenza di riconoscimento non può più essere riconosciuta e di conseguenza tagliata dalle nucleasi. Quindi gli HEG determinano una conversione dall'eterozigotità all'omozigotità che è irreversibile e provocano un'eredità distorta. Quando l'homing avviene nella linea germinale di un individuo eterozigote la maggior parte della progenie eredita l'unità, sul cromosoma originale che trasporta il gene o sul cromosoma appena convertito. Dunque, se la procedura avviene in maniera efficiente, più del 90% della progenie riceve l'HEG, rispetto al 50% previsto dall'eredità mendeliana (Figura 13). Il tasso di attività delle nucleasi nella linea germinale e la probabilità che le DSB siano riparate mediante un meccanismo di HDR, piuttosto che un meccanismo di NHEJ, determinano l'esito di diffusione del gene drive. L'applicazione di questa strategia su insetti pericolosi per la salute umana, come le zanzare, prevede la progettazione di HEG capaci di colpire geni essenziali per la fertilità femminile [41]. Nel caso delle *A.gambiae* sono stati condotti esperimenti in gabbia che hanno prodotto risultati positivi per il controllo della malaria e della dengue [52].

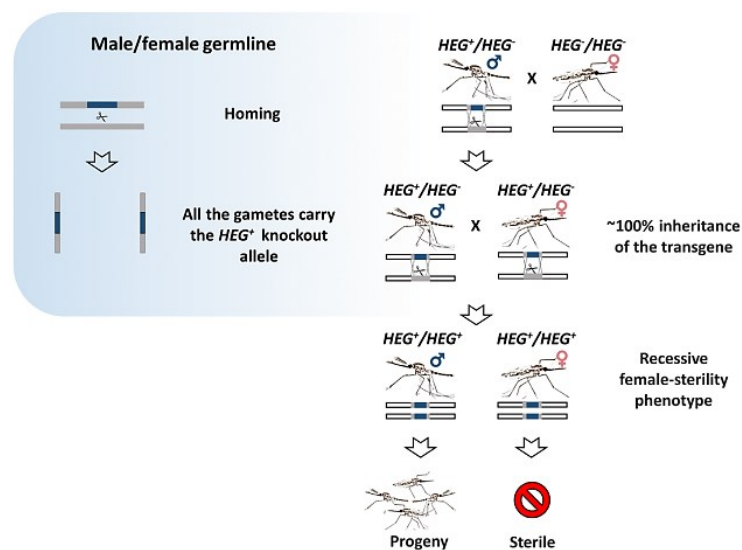


Figura 13: Sistema gene drive basato su HEG nella linea germinale che agisce su geni coinvolti nella fertilità femminile [41].

In un sistema gene drive, oltre alle HEG, possono essere utilizzate anche altre nucleasi, con una sequenza di riconoscimento sufficientemente lunga, come le *Zinc-finger nucleases* (ZFN) o le *Transcription activator-like effector nucleases* (TALEN). Sebbene siano facili da progettare, lo svantaggio principale è l'elevata tendenza a ricombinarsi durante l'homing.

3.3.7 Gene drive basati su CRISPR/Cas9 (CCGD)

Una tecnologia innovativa ed efficiente per la modifica genomica è il sistema CRISPR/Cas9. Il CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) è un insieme di sequenze di DNA che i batteri utilizzano come un vero e proprio sistema immunitario adattativo per proteggersi dai fagi tramite la scissione del DNA guidata dall'RNA, mentre le Cas9 rappresentano una classe di endonucleasi. L'editing genomico mediante CRISPR/Cas9 si basa su due elementi fondamentali:

- Una sequenza di RNA guida (gRNA): sequenza nucleotidica, facilmente programmabile, di circa 17-20 bp per dirigere l'attività delle endonucleasi nel sito target;
- Una proteina Cas9: endonucleasi associata a CRISPR che va ad agire sul sito bersaglio individuato dal gRNA.

L'attività della Cas9 dipende da una sequenza di 3-6 bp che si trova a valle del sito bersaglio ed è definita motivo protospaziatore adiacente (PAM). Il gRNA lega la Cas9 e la guida verso il sito target del genoma ospite, dove avviene la rottura del DNA a doppio filamento [52] [51]. Il processo di riparazione del DSB utilizza il cromosoma contenente il gRNA e la Cas9 come modello, attuando un meccanismo HDR e consentendo l'introduzione del tratto desiderato in un sito specifico del genoma (Figura 14). La possibilità di ingegnerizzare specifici gRNA permette, quindi, l'inserimento di sequenze desiderate affiancate da segmenti omologhi a quelli riconosciuti da Cas9 [45].

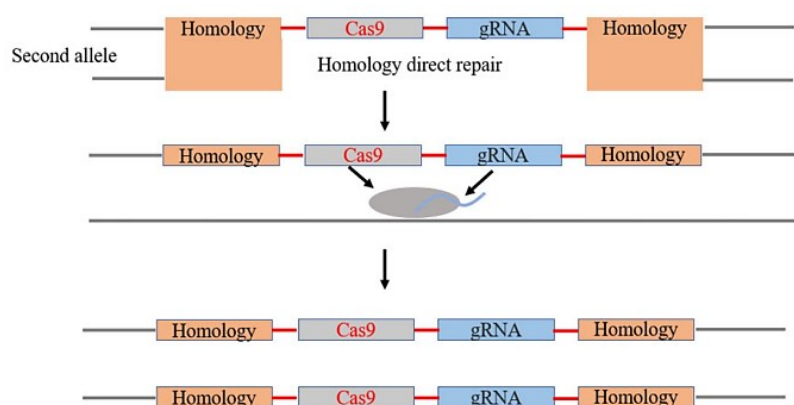


Figura 14: Sistema gene drive basato su CRISPR/Cas9 e meccanismo di HDR [51].

Il meccanismo HDR consente di replicare (oltre a Cas9 e gRNA) anche qualsiasi altro gene posto accanto alla Cas9, permettendo, dunque, al sistema il trasporto di un transgene che contiene un tratto desiderato. Al contrario, il secondo meccanismo di riparazione del DSB, il NHEJ porta all'eliminazione del gene.

È possibile distinguere due gruppi di sistemi gene drive CRISPR/Cas9:

- *Full gene drive* (fGD): in cui gRNA e Cas9 sono nella stessa cassetta¹⁵ e definito anche “gene drive collegato” (Figura 15a).
- *Split gene drive* (sGD): in cui gRNA e Cas9 sono in cassette diverse posizionate in diverse posizioni del genoma. Il gRNA è posizionato nel sito in cui la Cas9 taglia il DNA e la Cas9 è posizionata altrove nel genoma. La cassetta contenente il gRNA è funzionale solo se presente insieme alla cassetta con la Cas9. La cassetta Cas9 è ereditata in modo mendeliano, mentre la cassetta gRNA è copiata nel genoma. L'assenza della cassetta Cas9 implica un'ereditarietà mendeliana anche per la cassetta gRNA. Il sistema sGD consente un maggiore controllo del sistema e la possibilità di valutare gli effetti di ciascun componente (Figura 15b). Un particolare tipo di sGD è un sistema definito trans-complementante in cui è presente un secondo gRNA (gRNA2) che consente di dirigere la scissione nel sito di inserimento della Cas9, consentendo di copiare entrambe le cassette nel genoma (Figura 15c) [52].

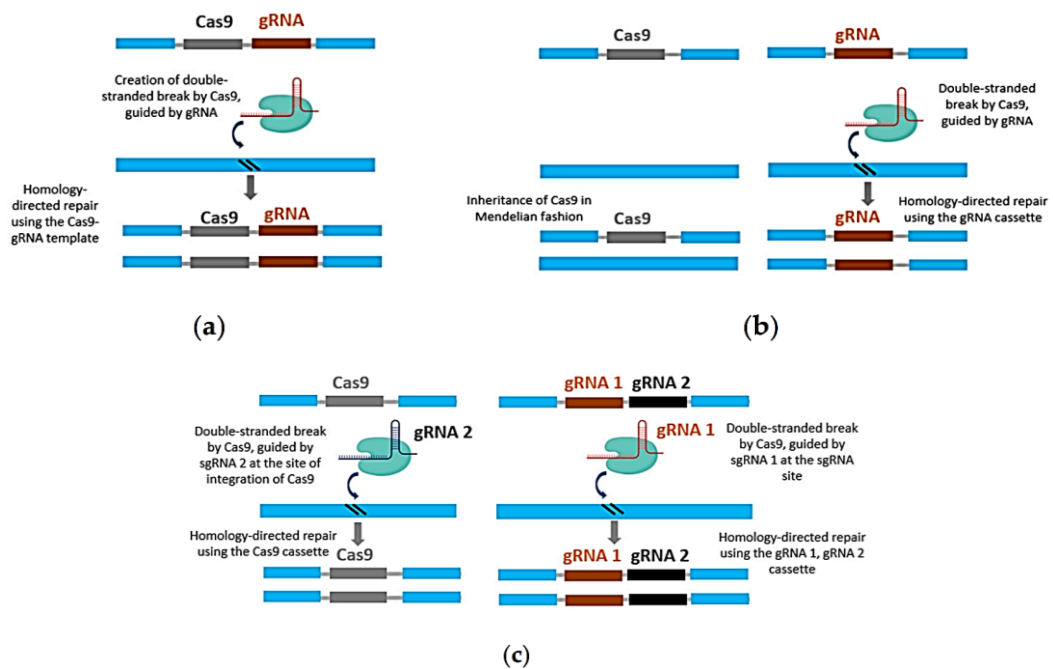


Figura 15: Sistemi gene drive basati su CRISPR/Cas9. Differenze tra sistema *full gene drive* (a), *split gene drive* (b) e gene drive *trans-complementante* [52].

¹⁵ Cassetta genica: elemento genico mobile che contiene un gene e un sito di ricombinazione.

Il sistema gene drive basato su CRISPR/Cas9, definito anche gene drive sintetico, solitamente include:

- Il gene che si intende inserire;
- Un gene marcatore, ad esempio, un gene che codifica per la proteina fluorescente rossa, DsRed;
- Uno o più gRNA;
- Un gene Cas9 controllato da un promotore della linea germinale (ad esempio *vasa*, *nanos* o *zpg*);
- Sequenze omologhe per guidare l'HDR .

L'utilizzo di un sistema gene drive basato su CRISPR/Cas9 ha dimostrato un'efficienza maggiore rispetto ai sistemi basati su HEG. Un primo problema associato all'uso di endonucleasi è la generazione di un taglio incompleto e improprio. Nel caso del gene drive basato su HEG si è verificata un'efficienza di taglio pari al 72% rispetto a quella registrata dalle Cas9 compresa tra l'88% e il 99%. Inoltre, un meccanismo di HDR, anziché di NHEJ, è fondamentale per la trasmissione del tratto genico desiderato ed è noto che i tassi di HDR varino tra le diverse specie, gli stadi di sviluppo e i diversi tipi cellulari. Un aumento della frequenza di HDR si verifica quando la Cas9 è attiva nelle prime fasi dello sviluppo, in particolare nei tessuti della linea germinale [45] [51].

Nelle zanzare *A.gambiae* i sistemi CRISPR/Cas9 mostrano tassi di homing di circa il 100%. In questo caso specifico, un gene drive può essere legato ad un gene effetto desiderabile (un carico) o eliminare alcuni geni essenziali della zanzara, solitamente coinvolti nella vitalità femminile e/o responsabili della fertilità. Nel primo caso vengono incorporati geni effettori anti-*Plasmodium*; al contrario, nel secondo caso, il gene drive causa un fenotipo recessivo e man mano che si diffonde nelle generazioni successive, gli individui eterozigoti sono vitali e continuano a trasmettere il gene drive, mentre gli individui omozigoti non sono vitali o risultano sterili, portando ad una riduzione della capacità riproduttiva del vettore [43].

Una particolare caratteristica dei sistemi basati su CRISPR/Cas9 è che il loro utilizzo possa non seguire il principio dell'HEG, ma quello di altri sistemi gene drive, come il drive meiotico dei cromosomi sessuali o Medea. Esperimenti di laboratorio hanno dimostrato che l'utilizzo del sistema CRISPR in un sistema Y-drive ha permesso di agire su una sequenza conservata di DNA ribosomiale situata sul cromosoma X. La compromissione di questa sequenza ha portato ad un controllo della malaria grazie alla generazione di zanzare maschi con una percentuale compresa

tra l'86,1% e il 94,8%. Tuttavia, la difficoltà principale sta nella difficile espressione delle nucleasi sul cromosoma Y, probabilmente a causa del fenomeno dell'inattivazione meiotica dei cromosomi sessuali. Il meccanismo TARE (embrione recessivo antitossina-antidoto), basato sul sistema Medea, invece, è stato testato su *D. melanogaster* prendendo di mira un allele letale recessivo. In questo caso è stata osservata una formazione più lenta degli alleli resistenti e la possibilità di recuperare la sua funzione in caso di resistenza [48] [52].

3.4 Limiti e sfide del gene drive

Nonostante i vantaggi, i sistemi gene drive suscitano diverse preoccupazioni relative all'efficacia e alla sicurezza delle tecniche utilizzate. In particolare, emergono considerazioni in merito alla sostituzione della specie, alla generazione di alleli resistenti, allo sviluppo di mutazioni indesiderate, ai meccanismi di riparazione del DSB e all'influenza del tratto effettore.

Nel tentativo di soppressione è possibile che si verifichi la sostituzione della specie di zanzara eliminata con un'altra specie presente. Inoltre, potrebbe essere necessario applicare la soppressione ripetutamente. L'idea potrebbe essere quella di eliminare la malattia a livello locale, in zone vicine, e progressivamente arrivare all'eliminazione su scala globale. I meccanismi di modifica risulterebbero più vantaggiosi in quanto non generano nicchie vuote e determinano un impatto ecologico minore. Inoltre, un ulteriore svantaggio è la necessità di costruire cassette anti-patogene specifiche, ciascuna per ogni specie di parassita della malaria [48].

La generazione di alleli funzionali resistenti al meccanismo gene drive può portare ad una riduzione o inversione dell'effetto benefico. *L.A.gambiae* presenta polimorfismi a singolo nucleotide (SNP) ogni 2,2 basi, quindi, una possibile resistenza alle nucleasi porterebbe ad un'interruzione della reazione di homing. Strategie individuate per attenuare la resistenza del sito target potrebbero essere: la limitazione dell'espressione delle endonucleasi a tessuti germinali definiti, l'attenta analisi dei siti target che ridurrebbero i mutanti potenzialmente resistenti e il targeting di sequenze multiple. In quest'ultimo caso, i vari sistemi gene drive possono essere indirizzati a colpire diversi siti del genoma, cosicché la probabilità che si verifichi una resistenza a tutte sia ridotta. Una maggiore efficacia sarebbe garantita se i siti fossero dispersi su un solo gene, poiché la stretta vicinanza dei bersagli ridurrebbe il processo di ricombinazione per riunire sequenze resistenti su un solo cromosoma. Anche in questo caso, però, le strategie multiplex potrebbero indurre tossicità a causa dell'aumento dei danni al DNA e della dose di nucleasi, derivata dall'utilizzo di più cassette di espressione [54].

La generazione di mutazioni che possono verificarsi al di fuori del sito target potrebbero portare all'alterazione delle funzioni biologiche del vettore. Le mutazioni sono involontarie ed ereditabili e compromettono la diffusione del vettore. I fattori che influiscono sull'insorgenza delle mutazioni sono: la specificità, l'attività, la dose e la regolazione spazio-temporale delle nucleasi, la sequenza della sua regione bersaglio, la presenza e frequenza di regioni genomiche simili e l'accessibilità di questi siti alle nucleasi [54].

L'utilizzo di un meccanismo di riparazione NHEJ da parte delle cellule germinali porterebbe all'introduzione di mutazioni impedendo l'inserimento del costrutto. Di conseguenza, le mutazioni trasmesse alla prole diffondono la resistenza alla modifica con la possibilità di generare fenotipi che potrebbero facilitare la trasmissione di patogeni [55].

Per i sistemi progettati per diffondere un tratto effettore esiste il rischio che l'effettore sia perso lasciando l'unità vuota. Ad oggi, non sono stati individuati transgeni effettori che determinano una piena resistenza al *Plasmodium*. È necessaria, infatti, una completa conoscenza della biologia del tratto effettore, in quanto fenotipi non desiderati potrebbero determinare la possibile generazione di alterazioni irreversibili. Diversi studi documentano che effettori mirati agli stadi di sviluppo precoci nell'intestino medio della zanzara siano particolarmente efficaci [42].

3.5 Biosicurezza dei sistemi gene drive

L'applicazione dei sistemi gene drive è impegnativa e innovativa e in quanto tale porta la comunità scientifica, e non solo, ad una attenta valutazione dei rischi il cui obiettivo è quello di bilanciare i benefici attesi e le potenziali conseguenze associate. È importante considerare gli effetti, a breve e lungo termine, che il rilascio di zanzare geneticamente modificate (GMM), incluse quelle che si basano su un sistema gene drive (GMGDM), possa avere sul fronte ecologico e ambientale.

Dal punto di vista biologico e ingegneristico una prima valutazione riguarda la sicurezza, l'efficienza e l'efficacia della tecnica di ingegneria genetica utilizzata. La ricerca e gli esperimenti di laboratorio devono essere condotti da personale qualificato che utilizza sistemi di protezione come, ad esempio, contenitori a tripla nidificazione in stanze sicure e ad accesso limitato. Quando possibile, devono essere messi in atto livelli di sicurezza volti al contenimento molecolare, garantendo che solo le popolazioni in laboratorio siano suscettibili al drive, e che gli esperimenti siano condotti in luoghi che sono al di fuori dell'aria abitabile [54]. L'OMS prevede che la sperimentazione, dal laboratorio all'eventuale rilascio, sia suddivisa in quattro

fasi. La strategia proposta si basa su criteri “go” e “no-go” in cui il passaggio ad una fase successiva è consentito solo se tutti i criteri della fase precedente vengono rigorosamente rispettati. Le quattro fasi sono:

- *Fase 1*: comprende fasi di progresso dagli studi di laboratorio a gabbie controllate utilizzando un numero basso di individui;
- *Fase 2*: prevede l'utilizzo di gabbie all'aperto più grandi, effettuando test sul campo. Le gabbie sono fisicamente contenute ed ecologicamente simili all'area destinata al rilascio delle zanzare;
- *Fase 3*: progetta sperimentazioni a rilascio aperto che aumentano in dimensioni, durata e complessità e valuta condizioni di efficacia nella prevenzione di malattie e infezione;
- *Fase 4*: realizza un intervento di applicazione più ampia e di sanità pubblica con valutazione dell'efficacia e della sicurezza.

In ogni fase vengono analizzati i possibili danni associati a quella fase [55]. L'utilizzo di gabbie di dimensioni più grandi consente di riprodurre condizioni ambientali simili a quelli naturali, controllando temperatura, umidità e ciclo di luce. Inoltre, permette non solo di stimolare i comportamenti naturali delle zanzare, come l'accoppiamento, ma anche di valutare l'idoneità e la competitività della popolazione modificata in generazioni consecutive. La validazione delle previsioni modellistiche della dinamica del gene drive e la rilevazione di eventi, come resistenza o mutazioni fuori bersaglio, rappresentano le ultime valutazioni effettuate. Quest'ultime aiutano nella comprensione dell'eterogeneità genetica e nella definizione di strategie di rilascio e monitoraggio post-rilascio [54].

Un aspetto fondamentale è la minimizzazione del rischio di rilascio accidentale e diffusione nella popolazione selvatica, garantita da strategie di confinamento e contenimento. Il livello di confinamento ad una specie o a un sottoinsieme di specie imparentate, è assicurato, naturalmente, dai sistemi gene drive grazie alla loro specificità, in particolare quelli basati sull'utilizzo delle endonucleasi. Nel controllo della malaria, per esempio, i gene drive sviluppati sono in grado di colpire le specie che trasmettono la malattia senza agire su più di 3000 specie di zanzare che non trasmettono malattie umane. Tuttavia, la generazione di mutazioni potrebbe avere un effetto in popolazioni non bersaglio. Il possibile rischio di diffusione al di fuori dell'area prevista ha portato alla definizione di strategie che consentono di rallentare o invertire i gene drive. Nel caso di drive soppressivi che potrebbero sviluppare una resistenza del sito bersaglio, vengono sviluppati dei sistemi gene drive che impediscono che questa resistenza si verifichi naturalmente. Gli insetti vengono progettati per portare una resistenza sintetica e il

loro rilascio consente la creazione di una “zona cuscinetto” che protegge da un’ulteriore diffusione man mano che la popolazione transgenica viene estinta [54].

Il rilascio di GMGDM implica la considerazione di possibili conseguenze ecologiche. Per prima cosa dovrebbe essere effettuata un’analisi quantitativa delle nicchie¹⁶ della specie e la valutazione dell’influenza che il rilascio di GMM possa avere sulle dinamiche della popolazione bersaglio. I fattori da considerare sarebbero fattori biologici, ambientali e fisici. Un’altra considerazione riguarda la generazione di una concorrenza intra-competitiva o inter-competitiva. La concorrenza intra-competitiva avviene tra individui di una specie, mentre quella inter-competitiva si verifica all’interno di nicchie concorrenti e porta all’aumento di una determinata specie e alla riduzione dell’altra. La valutazione della concorrenza permette lo studio della dinamica della popolazione. La dinamica comprende delle indagini riguardo alla stabilità, densità, mantenimento, declino ed estinzione nel tempo della popolazione. Ad esempio, un modello di dinamica della popolazione di *A. gambiae* ha suggerito una relazione lineare tra il numero di riproduzioni delle zanzare e le precipitazioni stagionali. Pertanto, è essenziale la definizione di una modellazione ecologica che possa definire i possibili impatti della tecnologia sull’ambiente [55].

3.6 Regolamentazione e normativa dei sistemi gene drive

La scienza, come spesso succede, è in anticipo rispetto agli sforzi comunitari e alle autorità governative. È necessario che i GMGDM, in quanto strumenti di sanità pubblica, siano affidabili per la salute umana e l’ambiente e dimostrino un’efficacia epidemiologica nel ridurre la prevalenza di malaria. Alla fase di implementazione in strutture opportunamente regolamentate, segue il rilascio sistematico nell’ambiente e l’incrocio con la popolazione selvatica. La pratica prevede non solo l’autorizzazione e la supervisione delle autorità nazionali, ma anche il coinvolgimento delle comunità locali. La preoccupazione principale è che gli attuali sistemi normativi non siano adeguatamente preparati agli eventuali rischi dei sistemi gene drive.

Un altro aspetto di interesse è associato al costo. È prevista, infatti, una fase di approvazione del prodotto che presuppone l’esecuzione di test sul campo in ciascun paese in cui si intende rilasciare i GMGDM. Il produttore è tenuto a fornire informazioni per accertare la validità del prodotto e comprovare che il metodo di produzione ne abbia preservato l’integrità. Allo sviluppatore sono attribuite, quindi, una serie di responsabilità, nonché l’obbligo di mettere a punto meccanismi di monitoraggio e segnalazione di eventi avversi. In particolare, devono

¹⁶ Nicchie ecologiche: indica il ruolo di una specie all’interno di un ecosistema.

essere specificati i tipi di eventi avversi che si verificano, la loro durata, la frequenza, quali azioni si intende mettere in atto per contrastare il problema e in che misura il produttore ne è responsabile.

Sono da considerare anche movimenti transfrontalieri dei sistemi rilasciati, per cui le autorità sono tenute a comunicare i dettagli del rilascio. Il movimento dei GMGDM è considerato vantaggioso da un Paese ricevente che ha dato il consenso al rilascio. Nel caso in cui il prodotto si sposti o sia trasportato in un Paese che non ha ricevuto l'approvazione per la biosicurezza, è possibile l'adozione di misure volte a rimuovere l'organismo. Misure per la rimozione possono essere messe in atto, anche nel caso in cui il prodotto si sposti attraverso l'incrocio con la popolazione locale e la loro presenza sia considerata illegale rispetto alle leggi normative di quel Paese. Infine, sarebbe necessario prendere in considerazione come commercializzare e consegnare i GMGDM e valutare se normative e politiche esistenti possano essere applicate al nuovo prodotto [56].

Capitolo 4

Applicazione di strategie di controllo genetico in letteratura

L'applicazione dei sistemi gene drive implica una conoscenza della biologia riproduttiva degli organismi coinvolti. Gli animali hanno sviluppato differenti meccanismi per la determinazione del sesso. Negli insetti, ad esempio, un segnale primario attiva un gene centrale che induce una cascata di meccanismi molecolari i cui prodotti modulano la differenziazione dell'organismo in maschio o femmina. Il meccanismo di determinazione del sesso è stato risolto solo in *D. melanogaster*.

4.1 Meccanismi di determinazione sessuale in *D.melanogaster* e *A. gambiae*

D. melanogaster ha un sistema di determinazione del sesso cromosomico XX/XY, in cui la doppia dose di elementi legati all'X nelle femmine attiva la trascrizione produttiva di un gene chiamato *sex lethal (sxl)*. *Sxl* blocca il meccanismo di compensazione del dosaggio¹⁷ e regola lo splicing¹⁸ di *transformer (tra)* in una forma produttiva che, insieme a *Tra2*, il suo cofattore, promuove lo splicing di *doublesex (dsx)* nella sua isoforma femminile (*dsxF*). Nei maschi una singola dose di XSE (elemento segnale codificato in X) è sufficiente ad inibire la cascata molecolare che si verifica nelle femmine, portando all'attivazione del meccanismo di compensazione di dose. In questo caso, *tra* è in una forma non produttiva e si verifica lo splicing nelle forme maschili di *dsx (dsxM)* e *fruitless, fru (fruM)*. Le forme maschili e femminili di *dsx* regolano le morfologie sesso-specifiche.

In *A. gambiae* non tutti i componenti del pathway di determinazione del sesso sono caratterizzati. Tuttavia, un recente studio [57] ha dimostrato l'esistenza di un gene, *femaleless (fle)*, che nelle femmine reprime la compensazione del dosaggio, secondo un meccanismo ancora sconosciuto, e regola lo splicing di *dsx* in una forma produttiva (*dsxF*) e *fru* in una forma non produttiva (*fruF*). La presenza di *fle* genera delle domande sul meccanismo di tale processo. In altri insetti, infatti, lo splicing di *dsx* è promosso da *tra* e dal suo cofattore *tra2*, entrambi però assenti in *A.gambiae*. Dunque, non è ancora chiaro se *fle* richieda un cofattore per lo splicing; inoltre, tutte le molecole a monte che determinano il sesso restano da identificare. Ciò che è noto è che nei maschi vi è la presenza di un determinante sessuale: *Yob*. *Yob* è un gene

¹⁷ Insieme di meccanismi che regolano l'attività dei geni X-linked.

¹⁸ Meccanismo che prevede la rimozione di sequenze non codificanti di un gene (introni) dall'RNA messaggero. Il processo di splicing porta alla generazione di un filamento di RNA maturo costituito esclusivamente da sequenze codificanti (esoni ed elementi di controllo).

situato sul cromosoma Y che avvia il processo di determinazione sessuale nei maschi inattivando *fle*. Come risultato dell'inattività di *fle* nei maschi il meccanismo di compensazione di dose è attivato e le isoforme maschili di *dsx* e *fru* vengono prodotte. (Figura 16) [57] [58].

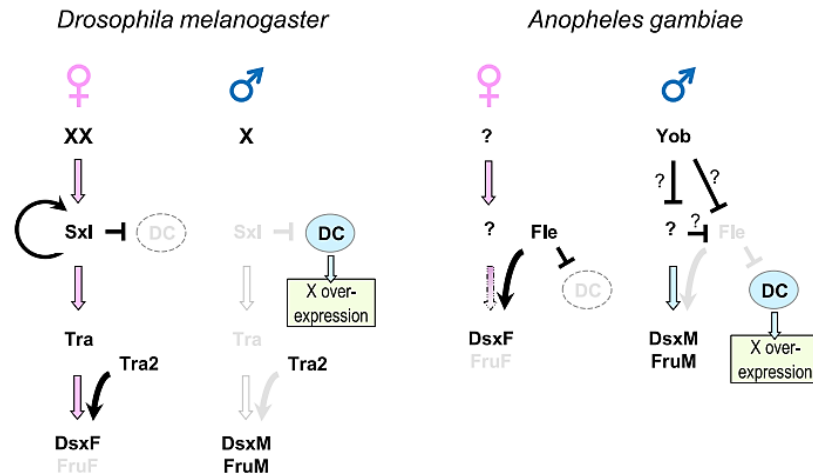


Figura 16: Meccanismo di determinazione sessuale in *D. Melanogaster* e *A. gambiae* dove DC indica il meccanismo di compensazione di dose. [57]

4.2 Doublesex e Yob in *A. gambiae*

Il gene *dsx* in *A.gambiae* (*Agdsx*) consiste di 7 esoni distribuiti su una regione di 85 kb sul cromosoma 2R. Il trascritto femminile, rispetto a quello maschile, presenta un esone 5 specifico per le femmine ed è altamente conservato (Figura 17) [53]. *Yob*, invece, si trova solo sul cromosoma Y dei maschi, è espresso negli embrioni maschili entro 2 ore dalla deposizione delle uova e la sua espressione è mantenuta durante tutta la vita della zanzara. L'attivazione di *Yob* precede l'inizio dello splicing sesso-specifico di *dsx*. L'isoforma femminile di *dsx*, depositata dalla madre, viene in gran parte degradata negli embrioni maschili entro 4 ore dalla deposizione delle uova; dopo la completa degradazione dell'isoforma femminile nei maschi è stabilito lo splicing di *dsx* specifico per il sesso [58].

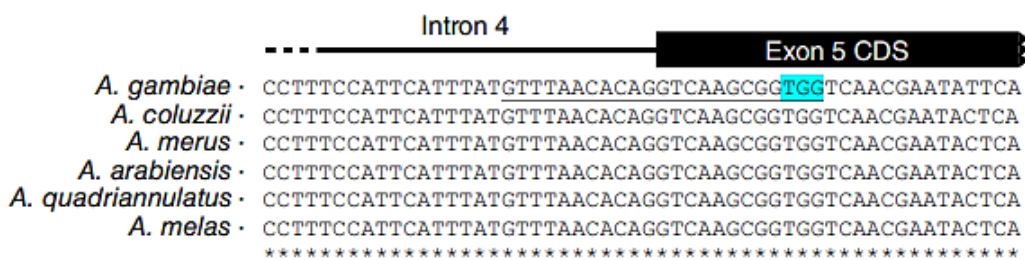


Figura 17: Rappresentazione del trascritto femminile *dsx* al confine introne 4-esone 5. L'allineamento in sei specie *Anopheles* dimostra che la sequenza è altamente conservata [53].

4.3 Studi

L'applicazione di un sistema gene drive ha come obiettivo la soppressione e/o l'eliminazione di insetti nocivi e vettori di malattie. A seguito della parziale comprensione dei meccanismi di determinazione sessuale in *A.gambiae* sono state proposte varie strategie per il controllo genico del vettore, alcune delle quali mirano ai diversi geni coinvolti nella determinazione del sesso.

4.3.1 Gene drive basato su CRISPR/Cas9 mirato a *dsx*

La volontà di voler sopprimere la capacità riproduttiva della popolazione di vettori della malaria ha portato alla conduzione di uno studio che in cui è stato interrotto il confine introne4-esone 5 di *dsx*, così da bloccare la formazione di *AgDsxF*. L'utilizzo di un RNA a singola guida (sgRNA) con una fonte di Cas9 e un modello per consentire l'HDR, iniettati in embrioni, ha portato alla generazione di individui eterozigoti trasformati ($dsxF^{CRISPRh/+}$) che sono stati incrociati con zanzare selvatiche. Le larve eterozigoti per la modifica dell'esone 5 si sono sviluppate in maschi e femmine adulti e in entrambi in casi è stata osservata un'elevata trasmissione di $dsxF^{CRISPRh}$ e nessun effetto a livello fenotipico. Al contrario, nel caso delle larve omozigoti ($dsxF^{-/-}$), una metà si è sviluppata in maschi normali, l'altra, invece, mostrava caratteristiche morfologiche sia maschili che femminili e una serie di anomalie degli organi riproduttivi (fenotipo intersessuale). Queste zanzare con un fenotipo intermedio si sono rivelate essere tutte cromosomicamente femmine (XX) e verranno da qui in avanti denominate intersex. Le zanzare intersex XX $dsxF^{-/-}$ risultavano inoltre sterili e incapaci di mordere [53].

Un altro studio mirato a colpire *dsx* è stato condotto definendo un sistema gene drive di distorsione sessuale (SDGD). Il sistema era basato su CRISPR e un'endonucleasi I-PpoI, capace di scindere una sequenza conservata nel pool di geni ribosomiali legati al cromosoma X e quindi causare le degradazioni del cromosoma X. Il costrutto $SDGD^{dsx}$ conteneva l'unità di

trascrizione *zpg-Cas9*, *beta2-IPpoI* e un gRNA; sono stati condotti quattro esperimenti in gabbia ed è stata osservata una distorsione sessuale a favore dei maschi e il collasso della popolazione. La distruzione dell'isoforma femminile di *dsx* e la conseguente riduzione della popolazione, diminuisce la probabilità che possano verificarsi mutazioni resistenti. Un aspetto fondamentale è che la riduzione della popolazione si è verificato a partire da una frequenza allelica del 2,5%. Questo risultato è compatibile con il numero di zanzare che si è capaci di produrre per esperimenti sul campo. Il sistema *SDGD^{dsx}* risulta più robusto di un normale sistema gene drive in relazione agli eventi che potrebbero verificarsi, infatti:

- l'eventuale perdita o inattivazione della sequenza I-PpoI comporterebbe la generazione di un gene drive funzionale *dsx* che contribuirebbe alla soppressione della popolazione;
- la perdita della funzione delle Cas9 o del gRNA produrrebbe alleli *dsx* non funzionali, tuttavia, la produzione di una progenie esclusivamente maschile sarebbe mantenuta grazie alla presenza di I-PpoI;
- le mutazioni che coinvolgono sia l'unità che il distorsore generano mutazioni *dsx* non funzionali. In quanto localizzate nel sito bersaglio delle nucleasi, le mutazioni non vengono selezionate perché non ripristinano la funzione del gene *dsx* e pertanto vengono perse [59].

4.3.2 Gene drive mirato a *Yob*

L'idea di rilasciare solo zanzare di sesso maschile ha portato alla sperimentazione di sistemi gene drive mirati all'utilizzo del gene *Yob*, tipico della popolazione maschile. È stato osservato che l'iniezione di *Yob* in embrioni di età inferiore a 2 ore porta allo sviluppo esclusivamente di zanzare maschi, poiché questo gene risulta essere letale per gli embrioni femminili. Lo studio è stato condotto iniettando anche un plasmide con una cassetta che presentava una proteina fluorescente verde (GFP) per monitorare il risultato dell'esperimento; in quanto la GFP è espressa transitoriamente dalle larve sviluppate dagli embrioni che hanno ricevuto sufficienti acidi nucleici. Allo stadio pupale è stato possibile osservare il sesso delle zanzare facendo un confronto non solo con un gruppo di controllo ma distinguendo anche tra gruppo GFP positivo (GFP+) e gruppo GFP negativo (GFP-). Tutti gli individui positivi alla GFP si sono sviluppati in maschi, mentre non è stato osservato un differente rapporto tra i sessi nel gruppo GFP negativo e nel gruppo di controllo (Figura18). Sebbene siano stati osservati risultati soddisfacenti, ad oggi, non esistono dei metodi in grado di generare l'elevato numero di zanzare maschili necessarie per il rilascio [58].

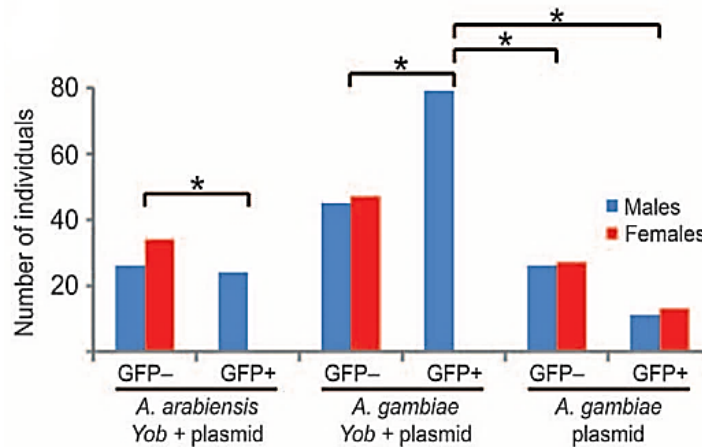


Figura 18: Risultati osservato dopo l'iniezione di Yob [58].

4.3.3 Elementi regolatori della linea germinale

Il processo di spermatogenesi porta alla formazione di spermatozoi altamente specializzati nelle gonadi maschili. Lo studio della linea germinale ha posto le basi per lo sviluppo di sistemi di controllo genico mirati ad indurre sterilità. Durante il processo di spermatogenesi, la regolazione dell'espressione genica risulta essere diversa tra i cromosomi sessuali e i cromosomi autosomici a causa di due meccanismi: la compensazione del dosaggio e l'inattivazione meiotica dei cromosomi sessuali (MSCI). La compensazione del dosaggio prevede una serie di meccanismi che regolano l'attività dei geni legati al cromosoma X. L'attività dei geni X-linked è infatti controllata da tre meccanismi principali: inattivazione di uno dei due cromosomi X negli organismi XX, iperattivazione del cromosoma X nell'organismo XY e ipoattivazione di entrambi i cromosomi X su un organismo XX. In *A.gambiae*, la compensazione del dosaggio è presente nelle pupe maschi, ad eccezione della linea germinale. Inoltre, durante la fase meiotica della spermatogenesi, i geni sul cromosoma X vengono repressi con un meccanismo simile all'MSCI. Una maggiore comprensione di questo meccanismo potrebbe essere fondamentale nello studio dei geni che vengono silenziati o attivati in diverse fasi della spermatogenesi.

L'uso di sequenze regolatrici della linea germinale per guidare l'espressione del transgene è fondamentale; tuttavia, solo alcuni di questi elementi regolatori sono stati caratterizzati e la maggior parte non mostra un'espressione precisa. Sia i sistemi gene drive basati sull'homing, sia i meccanismi Y-drive si basano sull'utilizzo delle endonucleasi, le quali presentano l'inconveniente di una ridotta efficienza causata dall'attività non sufficientemente limitata alla linea germinale. Un taglio incompleto o improprio del sito target da parte delle endonucleasi

può far sì che si verifichi una NHEJ del DSB, portando quindi all'insorgenza di mutazione e alla riduzione del tasso di trasmissione del drive.

È necessario, quindi, avere geni ed elementi regolatori con un'espressione fortemente regolata nella linea germinale [60]. Nei sistemi gene drive basati sull'utilizzo delle endonucleasi, ad esempio, quando l'espressione delle Cas9 è attivata nella linea germinale maschile/femminile, il taglio del sito bersaglio sul cromosoma wild-type e la seguente HDR determinano un aumento della frequenza del transgene nella popolazione. Alcuni esempi di regioni regolatorie utilizzate nelle zanzare sono: le regioni regolatorie del gene della *β2-tubulina*, utilizzate per guidare l'espressione di marcatori specifici del testicolo, e le regioni non tradotte dei geni *vasa*, *nanos* e *zero population growth (zpg)*, utilizzate per dirigere l'espressione delle nucleasi Cas9. È stato dimostrato che la deposizione paterna di Cas9 non determina la formazione di alleli resistenti, al contrario, la deposizione materna di complessi Cas9/gRNA potrebbe ostacolare la HDR e, di conseguenza, potrebbe verificarsi l'insorgenza di alleli potenzialmente resistenti al drive [50].

Capitolo 5

Materiali e Metodi

5.1 Dataset

Il dataset su cui è stata svolta l'analisi bioinformatica è stato prodotto presso l'Imperial College di Londra ed è disponibile nel Sequence Read Archive (SRA) con accession ID "PRJEB35365". Si tratta di trascritti di RNA di *A. gambiae* sequenziati mediante lo strumento Illumina HiSeq 2500. Il processo di sequenziamento produce sequenze chiamate reads. In questo caso, si tratta di un sequenziamento paired-end in cui entrambe le estremità di un frammento di RNA sono sequenziate e questo permette di generare dati più accurati, facilitare le operazioni di allineamento e rilevare eventuali duplicazioni, inserzioni o delezioni. Le reads ottenute da entrambe le estremità vengono generalmente chiamate "forward read" e "reverse read". I dati analizzati sono frutto di un sequenziamento di trascritti di RNA estratti da quattro distinte linee cellulari dalle gonadi maschili di *A. gambiae*. Le quattro popolazioni sono: premeiotica (pop I), meiotica I (spermatociti primari - pop II), meiotica II (spermatociti secondari - pop III) e postmeiotica (pop IV). Per ogni popolazione sono disponibili 3 repliche (Tabella 1).

5.2 Analisi bioinformatica

L'analisi dei trascritti di RNA permette di valutare se e quanto un gene è espresso in un campione biologico. Il confronto dell'espressione genica tra diversi campioni e/o condizioni si definisce analisi di espressione genica differenziale. L'analisi bioinformatica dei dati RNA-seq condotta prevede:

- il controllo di qualità delle reads;
- il pre-processing delle reads per la rimozione degli adattatori, delle basi di scarsa qualità, delle reads troppo corte e di quelle spaiate;
- il controllo di qualità delle reads dopo il pre-processing;
- l'allineamento al genoma di riferimento;
- la quantificazione dei trascritti;
- l'analisi dell'espressione genica;
- l'analisi dell'espressione genica differenziale;
- l'analisi di arricchimento funzionale.

	Type	Library Layout	Instrument	Organism	Sample Name	Tissue
ERR3661590	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_I_1	Male germline
ERR3661591	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_I_2	Male germline
ERR3661592	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_I_3	Male germline
ERR3661593	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_II_1	Male germline
ERR3661594	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_II_2	Male germline
ERR3661595	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_II_3	Male germline
ERR3661596	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_III_1	Male germline
ERR3661597	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_III_2	Male germline
ERR3661598	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_III_3	Male germline
ERR3661599	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_IV_1	Male germline
ERR3661600	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_IV_2	Male germline
ERR3661601	RNA-Seq	PAIRED	Illumina	<i>A. gambiae</i>	pop_IV_3	Male germline

Tabella 1: Caratterizzazione del dataset su cui è stata svolta l'analisi bioinformatica. Le popolazioni premeiotica, meiotica I, meiotica II e postmeiotica sono indicate rispettivamente con il nome di pop I, pop II, pop III e pop IV.

5.2.1 Controllo di qualità delle reads

Il controllo di qualità (QC) delle reads eseguito mediante il tool FASTQC [61] permette di verificare la qualità dei dati prima dell'esecuzione delle analisi successive. FASTQC genera dei report contenenti una serie di statistiche del file analizzato. Il *Summary* riassume tutti i controlli effettuati, affiancati da codici colore che indicano la bontà dei risultati ottenuti. Un'ulteriore sezione in alto contiene varie informazioni quali ad esempio il nome del file, il numero totale di reads sequenziate, la lunghezza della sequenza e il contenuto %GC. Le statistiche riportate sono:

- per base sequence quality (qualità per sequenza di base);
- per sequence quality score (punteggi di qualità per sequenza);
- per base sequence content (contenuto per sequenze di base);
- per sequence GC content (contenuto GC per sequenza);
- per base N content (contenuto di N per base);
- sequence length distribution (distribuzione della lunghezza della sequenza);
- sequence duplication levels (livelli di duplicazione della sequenza);

- overrepresented sequence (sequenze sovra-rappresentate);
- adapter content (contenuto di adattatori).

5.2.2 Processamento delle reads

Il tool Trimmomatic [62] è stato utilizzato per l'eliminazione delle basi di scarsa qualità e delle reads troppo corte e/o spaiate e per la rimozione degli adattatori (adapters), ovvero sequenze oligonucleotidiche necessarie per l'adesione dei frammenti al supporto in fase di sequenziamento. Si tratta di sequenze di scarsa qualità, la cui presenza può comportare la generazione di risultati non ottimali nelle analisi successive. Gli adattatori sono legati ad entrambe le estremità delle reads paired-end e vengono definitivi rispettivamente come "Read 1 Adapter" e "Read 2 Adapter". In questa fase di pre-elaborazione è necessario, quindi, assicurarsi che le sequenze contaminanti vengano rimosse mantenendo intatti i dati della sequenza validi [62]. In caso di reads paired-end, come quelle prodotte per il dataset in analisi, al tool vengono forniti in input 2 files, uno per le forward reads e uno per le reverse reads, e sono generati in output 4 files: forward paired e reverse paired, che indicano le reads sopravvissute all'elaborazione, e forward unpaired e reverse unpaired, in cui solo una read è sopravvissuta (Figura 19).

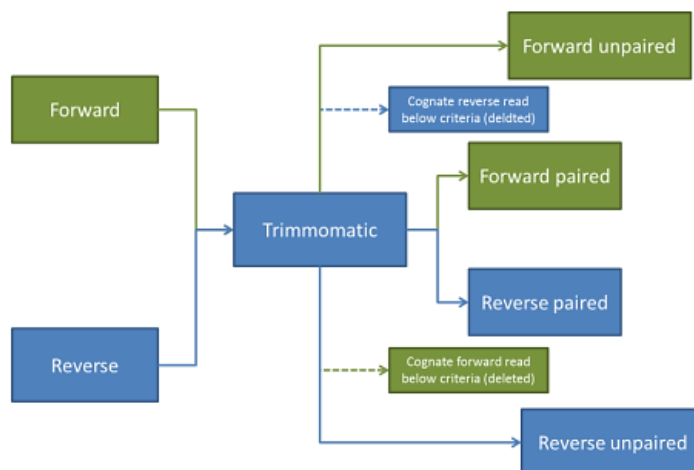


Figura 19: Input e output utilizzati dal tool Trimmomatic con dati paired-end.

Le operazioni eseguite da Trimmomatic vengono effettuate sulla base di una serie di parametri forniti dall'utente. Nel nostro caso i parametri utilizzati sono riportati nella tabella 2.

PARAMETRO	VALORE	SIGNIFICATO
PE	-	Indica il tipo di dati Paired-End
ILLUMINACLIP	{Path_adapters}:2:30:10	<ul style="list-style-type: none"> - Specifica il percorso del file contenente gli adattatori; - Specifica il numero massimo di mancate corrispondenze che consente comunque l'esecuzione di una corrispondenza completa; - Specifica quanto deve essere accurata la corrispondenza fra due letture; - Specifica l'accuratezza della corrispondenza tra qualsiasi sequenza di adattatori rispetto a una lettura.
LEADING	28	Rimuove le basi di bassa qualità dall'inizio e specifica la qualità minima richiesta per mantenere una base.
TRAILING	28	Rimuove le basi di bassa qualità dalla fine e specifica la qualità minima richiesta per mantenere una base.
SLIDINGWINDOW	4:28	Specifica il numero di basi di cui calcolare la media e la qualità media richiesta.
MINLEN	80	Specifica la lunghezza minima delle reads da mantenere.

Tabella 2: Parametri utilizzati in Trimmomatic.

5.2.3 Allineamento al genoma di riferimento

L'allineamento consente di mappare le reads al genoma di riferimento per determinare la posizione in cui una reads viene mappata. Il programma utilizzato per l'allineamento delle reads al genoma di riferimento di *A. gambiae* è HISAT2 [63]. Il genoma di riferimento utilizzato è VectorBase-62_AgambiaePEST_Genome.fasta disponibile in VectorBase [64]. HISAT2 è stato utilizzato con impostazioni standard senza specificare particolari parametri. Si tratta di un sistema altamente efficiente e rapido basato sull'utilizzo di due tipi di indici: un indice FM globale che rappresenta l'intero genoma e un insieme di piccoli indici FM per le regioni che coprono collettivamente l'intero genoma. I piccoli indici sono chiamati "indici locali" e combinati a diverse strategie di allineamento consentono una mappatura rapida e accurata. HISAT2 salva i risultati dell'allineamento in files in formato SAM (Sequence Alignment Map) o BAM (Binary Alignment Map). Per l'esplorazione visiva dei dati è stato utilizzato lo strumento interattivo IGV (Iterative Genomics Viewer) [65]. L'analisi e la manipolazione degli allineamenti sono state effettuate utilizzando SAMtools [66] [67]. Si tratta di un pacchetto che permette di lavorare sui dati di allineamento utilizzando vari comandi. Quelli utilizzati nell'analisi sono: *view* che visualizza, filtra e converte i file SAM/BAM/CRAM; *sort* che ordina i file SAM/BAM/CRAM; *idxstats*, *flagstat* e *stats* che permettono di calcolare varie statistiche sui file di allineamento; *markdup* che contrassegna gli allineamenti duplicati in un file ordinato per coordinate.

SAMtool viene utilizzato nuovamente in uno step successivo dell'analisi bioinformatica per rimuovere dal dataset le reads che mappano su geni mitocondriali.

5.2.4 Quantificazione dei trascritti

Il totale delle reads allineate su una regione di interesse del genoma di riferimento è detto counts. I risultati ottenuti dall'allineamento non sono sufficienti per l'interpretazione biologica; è necessario, infatti, riassumere i risultati in termini di copertura delle reads sul genoma di riferimento. L'operazione di quantificazione è stata effettuata con il tool featureCounts [68]. Oltre ai parametri standard sono stati utilizzati ulteriori parametri riportati nella tabella 3.

PARAMETRO	SIGNIFICATO
T	Specifica il numero di threads. In questo caso impostato ad 8
P	Conta le coppie di reads anziché le singole reads
B	Conta solo le coppie di reads che hanno entrambe le estremità allineate correttamente
P	Controlla la validità della distanza accoppiata durante il conteggio delle coppie lette
C	Non conta le reads lette le cui due estremità sono mappate su cromosomi diversi o mappate sullo stesso cromosoma ma su filamenti diversi
J	Conta in numero di reads che supportano ciascuna giunzione esone-esone.

Tabella 3: Parametri aggiuntivi utilizzati in featureCounts.

I dati di input di featureCounts consistono in uno o più file di reads allineate, in formato SAM o BAM, e un elenco di caratteristiche genomiche. I conteggi grezzi delle reads effettuati da featureCounts non possono essere utilizzati direttamente per il confronto dei livelli di espressione tra i campioni; pertanto, vengono utilizzate delle misure che permettono di tenere conto di fattori che influenzano la quantificazione, quali ad esempio le differenze nella lunghezza della trascrizione, la profondità di sequenziamento e la variabilità tra i campioni analizzati. La misura utilizzata nell'analisi si basa sui TPMs (Transcripts per Million). I TPMs sono calcolati come:

$$TPM = \frac{A}{\sum(A)} \times 10^6 \quad \text{con} \quad A = \frac{\text{total reads mapped to gene} \times 10^3}{\text{gene length in bp}}$$

L'operazione di quantificazione dei trascritti attraverso featureCounts è stata eseguita nuovamente dopo la rimozione dei geni mitocondriali. Anche in questo caso sono stati calcolati

i TPMs per ogni gene. Prima di procedere all'analisi dell'espressione genica viene eseguita una PCA implementata in R, sia prima che dopo la rimozione del genoma mitocondriale.

5.2.5 Analisi dell'espressione genica

La visualizzazione dell'espressione genica è stata eseguita utilizzando il pacchetto ggplot2 [69] di R. Si è deciso di rappresentare i livelli di espressione di particolari geni coinvolti nel processo di determinazione del sesso e di specifici promotori in *A. gambiae*. I geni rappresentati sono *doublesex* (AGAP004050), *femaleless* (AGAP013051) e *yob* (AGAP029221), mentre i promotori selezionati sono *exu* (AGAP007365), *nanos* (AGAP006098), *vasa2* (AGAP008578), *zpg* (AGAP006241).

L'espressione genica è stata ricalcolata nuovamente per gli stessi geni e promotori dopo aver rimosso il genoma mitocondriale.

5.2.6 Analisi dell'espressione genica differenziale

L'analisi dell'espressione genica differenziale è stata eseguita utilizzando il pacchetto edgeR [70] implementato in R. Dopo aver elaborato i dati e valutato le stime di dispersione, sono stati eseguiti dei test per la definizione dei geni up-regulated e down-regulated. EdgeR è supportato solo per confronti a coppie, per cui considerando il dataset in analisi composto da quattro popolazioni rispettivamente premeiotica, meiotica I, meiotica II e postmeiotica, il confronto è stato effettuato tra: premeiotica vs meiotica I, premeiotica vs meiotica II, premeiotica vs postmeiotica, meiotica I vs meiotica II, meiotica I vs postmeiotica e meiotica II vs postmeiotica.

Prima di effettuare i confronti tra le varie popolazioni, edgeR prevede una serie di steps. Per prima cosa è richiesta una fase di filtraggio in cui vengono mantenuti i geni che hanno un conteggio minimo di reads in un numero significativo di campioni. Il filtraggio è basato sui CPM (Counts per Million), ovvero una misura che permette di normalizzare i dati per la profondità di sequenziamento ma non per la lunghezza del gene. Il numero minimo di reads è definito in base alle dimensioni della libreria, come mostrato in figura 20.

Library size	Count	CPM
1M	1	1
10M	10	1
20M	10	0.5

Figura 20: Relazioni tra dimensioni della libreria, count e CPM.

Nel nostro caso la dimensione della libreria è diversa per ogni campione; perciò, è stata calcolata una media in maniera tale da poter scegliere una soglia appropriata. La media era pari a 12.610.843, di conseguenza è stato impostato un valore di minimo di count pari a 10.

Successivamente, come suggerito in [71], è stato eseguito un passaggio che ha permesso di normalizzare le dimensioni della libreria, così da minimizzare i cambiamenti nel log-Fold. Il metodo di normalizzazione utilizza la TMM (Trimmed Mean of M-value) basato sull'ipotesi che la maggior parte dei geni non sono differenzialmente espressi.

Prima di eseguire l'analisi di espressione genica differenziale, è obbligatorio definire se l'esperimento è a singolo fattore o a fattori multipli. Sebbene l'unica variabile nel nostro caso sia il tempo, essa va considerata nei quattro diversi tipi di popolazione cellulare. Lo scorrere del tempo determina una differenza tra le quattro popolazioni, per cui si tiene conto sia del tipo di popolazione cellulare sia delle variazioni nel tempo all'interno di ciascuna popolazione. Date queste considerazioni si è scelto di considerare l'esperimento come un esperimento a fattori multipli. La scelta di questo design sperimentale influisce sulla stima della dispersione. In questo caso, infatti, la stima della dispersione è calcolata con il metodo della verosimiglianza aggiustata per i profili di Cox-Reid (CR) basato sull'utilizzo di modelli lineari generalizzati (GLM). I GLM sono una classe di modelli che derivano da quelli lineari classici e permettono di modellare l'espressione genica in relazione ai diversi fattori che influiscono sulla variabile di risposta. Il GLM è più flessibile rispetto ad un modello lineare classico poiché permette di gestire diverse distribuzioni non necessariamente normali come quelle assunte dal modello lineare classico. Inoltre, il modello classico presuppone un legame lineare tra la variabile di risposta e le covariate (fattori), al contrario, nel GLM è possibile specificare una serie di funzioni di legame a seconda della distribuzione della risposta e della relazione tra la risposta e i fattori. Utilizzando il GLM vengono stimati tre tipi di dispersione: una 'common dispersion', una 'trended dispersion' e una 'tagwise dispersion'. La 'common dispersion' calcola un valore di dispersione comune per tutto il set di dati, quindi, in questi casi si assume che tutti i geni

abbiano la stessa relazione media-varianza. La ‘trended dispersion’ indica la dispersione media tra tutti i geni con abbondanza simile; al contrario, la ‘tagwise dispersion’ stima la dispersione di ciascun gene.

A questo punto per la determinazione dell’espressione genica differenziale è possibile utilizzare due tipi di test: il likelihood ratio test (LRT) o il quasi-likelihood (QL) F-test. Al contrario di quanto fatto in [72] in cui è stato applicato il LRT, nel nostro caso è stato utilizzato il QL F-test in quanto fornisce un controllo del tasso di errore più robusto e affidabile. Il QL F-test fornisce una serie di statistiche delle espressioni differenziali:

- log FC: il Fold Change è utilizzato per valutare il cambiamento di espressione dei geni nelle due condizioni. Un fold change positivo indica un aumento di espressione, mentre un cambiamento negativo indica una diminuzione;
- log CPM: la media del logaritmo dei counts per million;
- log F: indica la statistica F di quasi-likelihood per il GLM;
- P-value: indica la probabilità che il gene analizzato venga espresso in modo differenziale;
- FDR: False Discovery Rate permette di minimizzare l’errore di tipo II e tenere sotto controllo l’errore di tipo I. l’errore di tipo II consiste nell’accettare l’ipotesi nulla quando essa in realtà è falsa, mentre l’errore di tipo I è la probabilità di rifiutare un’ipotesi nulla quando essa è vera. Nel nostro caso si parte da un’ipotesi nulla in cui i geni non sono differenzialmente espressi.

Di tutti i geni differenzialmente espressi ottenuti dai confronti a coppie, sono stati tenuti solo quelli con un $\log_2(\text{FC}) > |1|$. La soglia è stata scelta perché utilizzando un \log_2 si considera un Fold Change che sia almeno il doppio nelle due condizioni. Successivamente è stata applicata un’ulteriore soglia tenendo i geni con un $\text{FDR} < 0.05$.

5.2.7 Analisi dell’arricchimento funzionale

L’analisi dell’arricchimento funzionale permette l’identificazione di specifici processi biologici che descrivono le attività molecolari o i ruoli dei geni. Gene Ontology (GO) [73] è un’ontologia comunemente utilizzata per la descrizione dell’annotazione funzionale dei geni. Il sistema di classificazione PANTHER [74] è stato utilizzato per classificare e identificare la funzione dei prodotti genici. L’intera analisi è stata eseguita utilizzando il pacchetto `rbioapi`, implementato in R, e applicando il test esatto di Fisher con correzione del FDR. Lo studio è stato effettuato

utilizzando la funzione `rba_panther_enrich` all'interno della quale sono stati passati i seguenti parametri:

- l'id dei geni da confrontare;
- l'id dell'organismo considerato (*A. gambiae*=7165);
- l'id delle ontologie da indagare. Nel nostro caso sono state utilizzate le ontologie molecular function (GO:0003674), biological process (GO:0008150) e cellular component (GO:0005575);
- il tipo di test statistico per calcolare i p-values;
- metodo di correzione.

L'analisi di arricchimento funzionale è stata eseguita per ogni gruppo di geni up-regulated e down-regulated ottenuto dai confronti a coppie effettuati nella DGE. In genere i risultati dell'analisi sono disposti in una tabella di otto colonne con:

- number in list: indica numero di geni nell'elenco caricato che corrispondono alla categoria di dati di annotazione;
- fold enrichment: indica il rapporto tra il numero di gene nell'elenco caricato rispetto al numero previsto. Se è maggiore di 1 indica che la categoria è sovra-rappresentata, al contrario, se è minore di 1 la categoria è sotto-rappresentata.
- fdr: false discovery rate calcolato con la procedura di Benjamini-Hockberg. Per filtrare i risultati viene utilizzato un valore critico pari a 0.05.
- expected: indica il numero di geni che ci si aspetterebbe nell'elenco per questa categoria in base all'elenco di riferimento;
- number in reference: contiene il numero di geni nell'elenco di riferimento mappati alla categoria di dati di annotazione;
- p-value: indica la probabilità di osservare almeno un numero x di geni su un totale di n geni nell'elenco di annotazioni per un particolare termine GO. Più il valore p è vicino a zero, maggiore è la significatività di quel particolare termine GO associato a un gruppo di geni; questo significa che è meno probabile che l'annotazione osservata per il termine GO a un gruppo di geni si verifichi per caso;
- plus-minus: contiene un segno + o un segno -. Il segno + indica che questa categoria nell'esperimento è sovra-rappresentata e quindi sono stati osservati più geni del previsto in base all'elenco di riferimento. Il segno - indica una sottorappresentazione.
- term id: id della categoria di dati di annotazione;
- term label: il nome della categoria di dati di annotazione.

I risultati ottenuti in queste tabelle sono stati filtrati e sono stati tenuti i dati in cui si aveva un $\text{fold enrichment} > 2$, un $\text{fdr} < 0.05$ e una sovra-rappresentazione di quella particolare categoria.

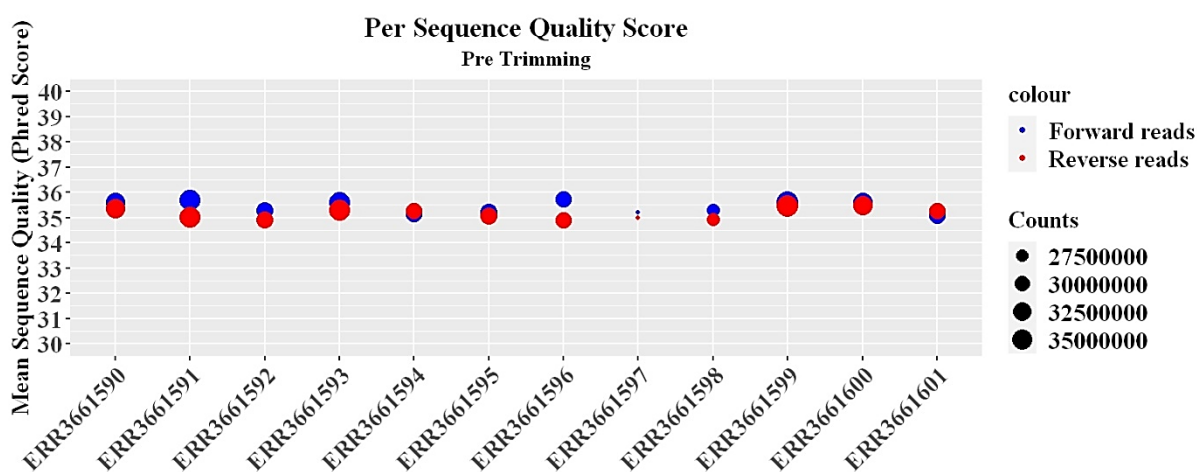
Capitolo 6

Risultati

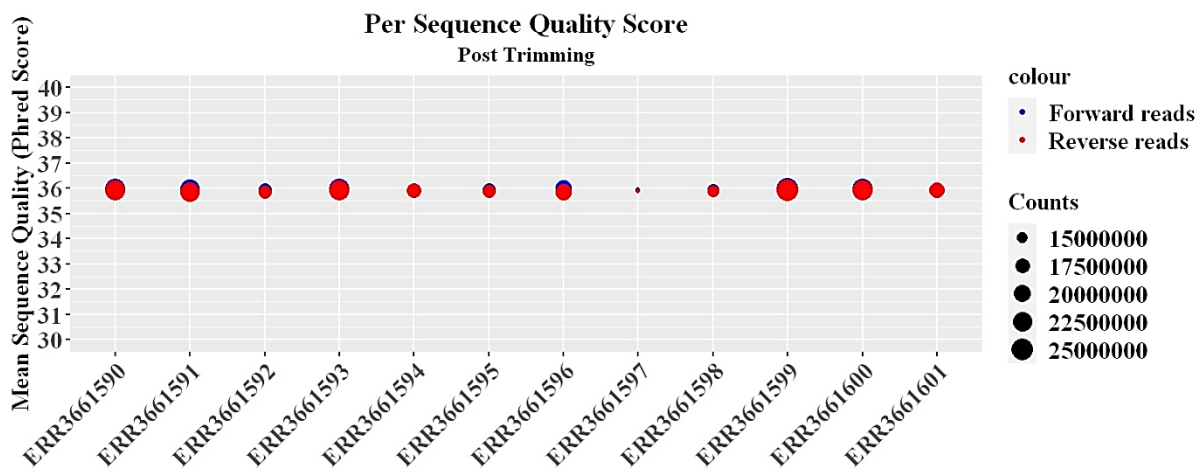
I risultati dei vari step dell'analisi bioinformatica condotta sui dati RNA-seq di *A. gambiae* sono stati investigati sia tramite l'ispezione visiva nel caso di grafici generati dai vari tool, sia attraverso la consultazione di tabelle create appositamente per consentire una più rapida interpretazione dei dati.

6.1 Quality Control

A causa dell'elevato numero di report generati dal tool FASTQC e l'impossibilità di effettuare un confronto immediato tra i dati, è stato necessario creare dei grafici in R che riportassero le statistiche dei campioni simultaneamente. Le statistiche analizzate sono: per sequence quality score e sequence length distribution. Entrambe sono state rappresentate prima e dopo il processamento. Le figure 21a e 21b riportano i punteggi di qualità prima e dopo l'operazione di trimming, sia per le forward reads che per le reverse reads. I risultati relativi alla distribuzione della lunghezza della sequenza vengono riportati nelle figure 22a e 22b.

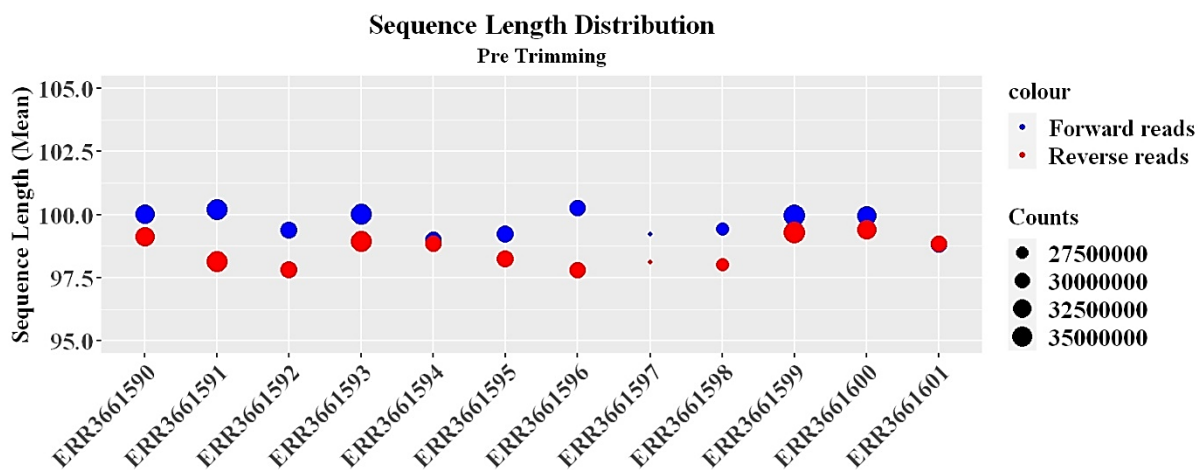


(a)

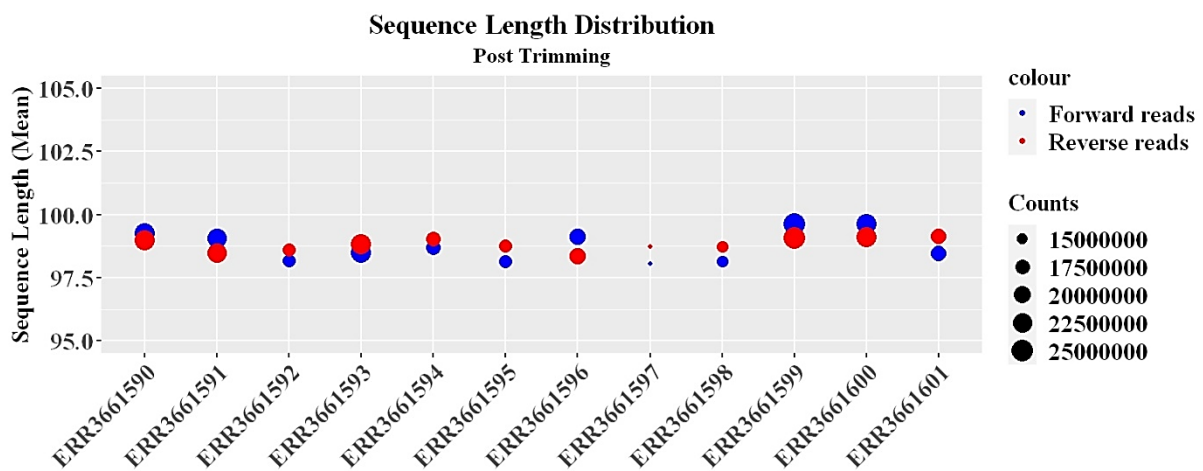


(b)

Figura 21: Plot dei punteggi di qualità relativi ai dati RNA-seq di *A. gambiae* prima del procesamiento (a) e dopo (b).



(a)



(b)

Figura 22: Plot della distribuzione della lunghezza delle sequenze relativa ai dati RNA-seq di *A. gambiae* prima del procesamiento (a) e dopo (b).

6.2 Allineamento al genoma di riferimento

Dopo aver effettuato l'allineamento, grazie all'applicazione di SAMtool, è stato possibile verificare il numero di duplicati e di reads correttamente accoppiate. La Tabella 4 mostra il numero di duplicati e il totale di reads per ogni campione, mentre la Figura 23 rappresenta la quantità di duplicati sul totale calcolata in percentuale. Nella Tabella 5 viene riportata la quantità e la percentuale di reads correttamente accoppiate per ogni campione. Nella Figura 24 la percentuale di reads accoppiate è rappresentata mediante grafici a barre.

Samples	Duplicates	Total
ERR3661590	6.789.595	42.748.514
ERR3661591	6.429.202	41.824.326
ERR3661592	4.076.806	30.167.630
ERR3661593	7.951.832	42.719.400
ERR3661594	5.675.439	31.741.349
ERR3661595	4.726.863	29.789.178
ERR3661596	6.735.256	34.899.190
ERR3661597	4.331.468	24.184.177
ERR3661598	4.897.394	27.642.920
ERR3661599	17.126.437	46.120.843
ERR3661600	15.248.442	42.664.844
ERR3661601	8.596.708	32.529.153

Tabella 4: Conteggio di duplicati e numero totale di reads per ogni campione.

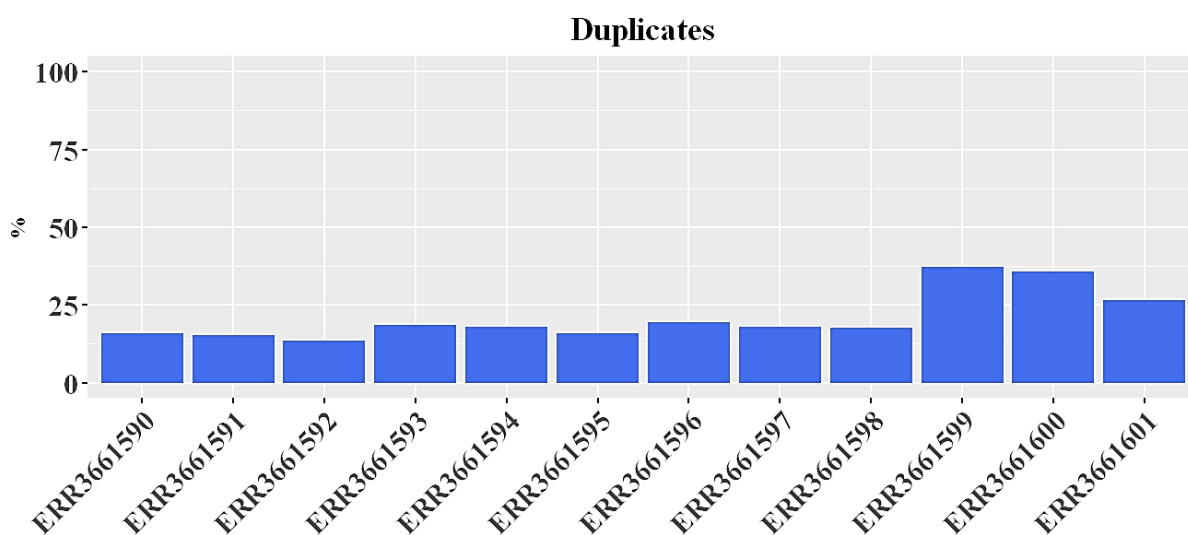


Figura 23: Percentuale di duplicati sul totale per ogni campione

Samples	Properly paired (%)	Reads
ERR3661590	98.40	39.741.832
ERR3661591	98.36	38.795.298
ERR3661592	97.95	27.995.122
ERR3661593	98.24	39.406.324
ERR3661594	98.53	29.507.982
ERR3661595	97.97	27.498.834
ERR3661596	98.49	32.099.462
ERR3661597	98.30	22.265.990
ERR3661598	98.05	25.398.068
ERR3661599	98.53	42.531.804
ERR3661600	98.55	39.377.294
ERR3661601	98.42	29.998.778

Tabella 5: Conteggio e percentuale di reads correttamente accoppiate per ogni campione.

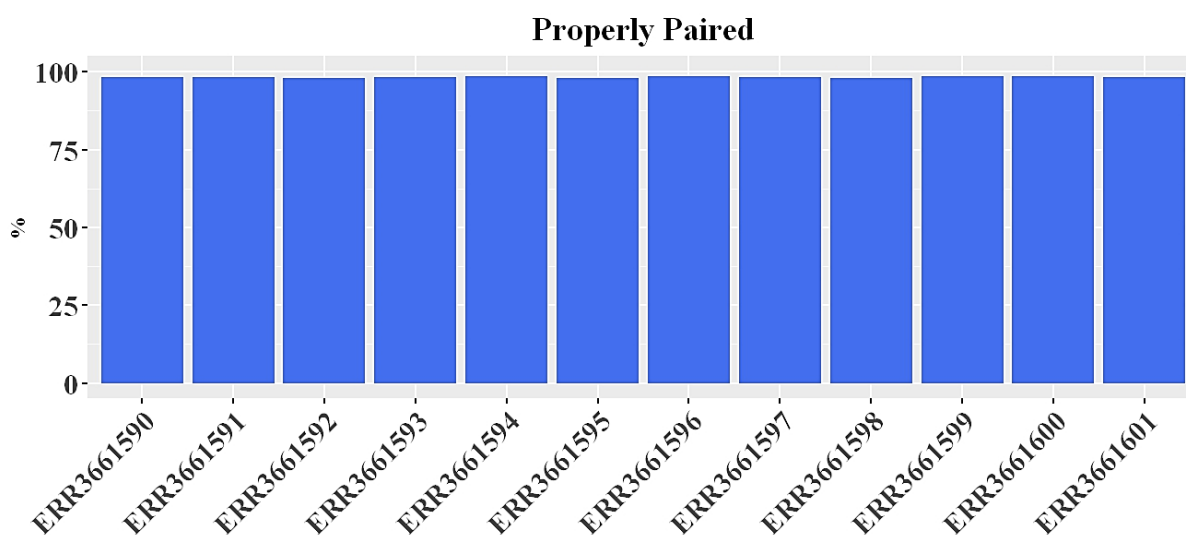


Figura 24: Rappresentazione delle reads correttamente accoppiate in percentuale per ogni campione.

Il contributo dei geni mitocondriali in ogni campione è riportato nella tabella 6.

Campioni	Mit_reads	Reads_tot	Mit_reads [%]
ERR3661590	233.465	17.553.956	1,329985
ERR3661591	170.922	17.144.897	0,996926
ERR3661592	213.074	12.292.773	1,733327
ERR3661593	176.192	17.295.563	1,018712
ERR3661594	112.895	12.948.374	0,871886
ERR3661595	134.513	11.889.626	1,131348
ERR3661596	51.405	13.662.338	0,376253
ERR3661597	37.731	9.556.385	0,394825
ERR3661598	56.560	10.720.485	0,527588
ERR3661599	193.839	17.936.605	1,080689
ERR3661600	149.070	16.878.223	0,883209
ERR3661601	105.856	12.887.480	0,821386

Tabella 6: Contributo dei geni mitocondriali sul totale delle reads per ogni campione.

Il conteggio dei duplicati e delle reads correttamente accoppiate viene effettuato di nuovo dopo la rimozione del genoma mitocondriale. Un confronto tra i dati prima e dopo l'eliminazione dei geni mitocondriali è riportato nella tabella 7.

Samples	Duplicates	Duplicates after filtering	Properly paired [%]	Properly paired after filtering [%]
ERR3661590	6.789.595	6.416.081	98.40	98.42
ERR3661591	6.429.202	6.178.667	98.36	98.38
ERR3661592	4.076.806	3.751.147	97.95	97.96
ERR3661593	7.951.832	7.720.824	98.24	98.27
ERR3661594	5.675.439	5.515.687	98.53	98.54
ERR3661595	4.726.863	4.546.736	97.97	98.00
ERR3661596	6.735.256	6.674.704	98.49	98.50
ERR3661597	4.331.468	4.295.138	98.30	98.30
ERR3661598	4.897.394	4.822.424	98.05	98.06
ERR3661599	17.126.437	16.810.058	98.53	98.53
ERR3661600	15.248.442	15.026.997	98.55	98.55
ERR3661601	8.596.708	8.449.988	98.42	98.42

Tabella 7: Confronto di duplicati e reads correttamente accoppiate prima e dopo la rimozione del genoma mitocondriale.

6.3 Quantificazione dei trascritti

La Principal Component Analysis (PCA) è stata eseguita sui TPMs e permette di proiettare il set di dati di grandi dimensioni su un grafico bidimensionale e raggruppare i campioni. La figura 25 mostra il grafico bidimensionale delle componenti principali che contribuiscono a spiegare la maggiore variabilità dei dati. In questo caso vengono rappresentate le prime due componenti principali.

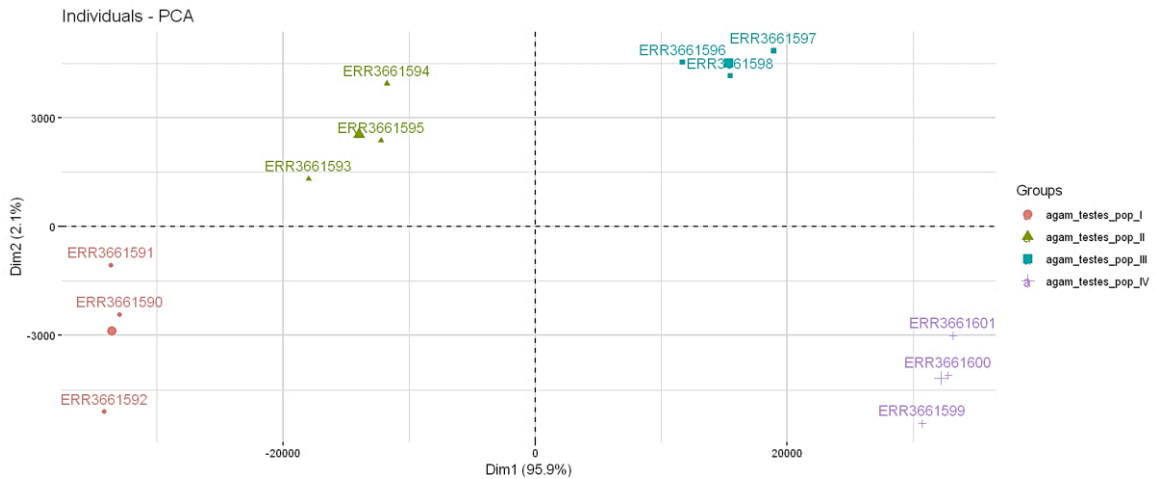


Figura 25: Analisi delle componenti principali (PCA) eseguita sui trascritti di *A. gambiae*.

La distribuzione dei TPMs è stata rappresentata attraverso dei boxplot nella figura 26.

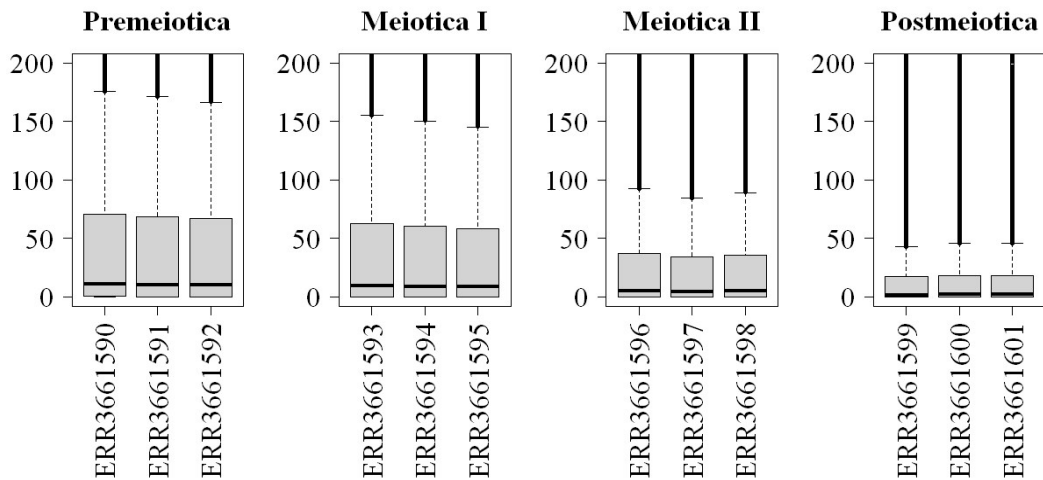


Figura 26: Distribuzione dei TPMs nei campioni suddivisi per popolazione.

In figura 27 vengono riportati i risultati della PCA dopo la rimozione dei geni mitocondriali per verificare che la corretta distinzione delle quattro popolazioni sia mantenuta.

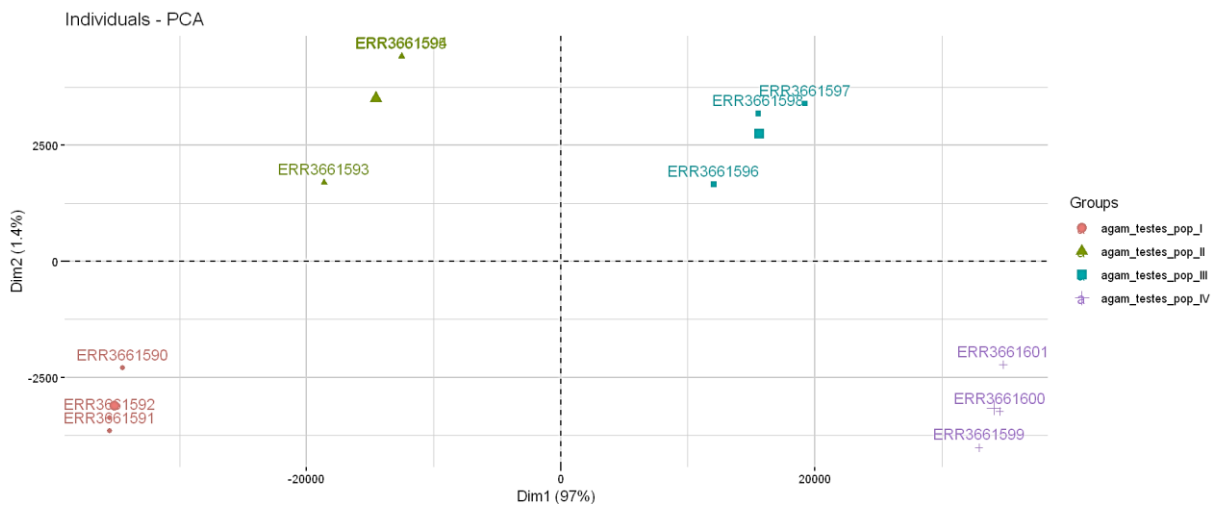


Figura 27: PCA eseguita sui trascritti di *A. gambiae* dopo la rimozione del genoma mitocondriale.

Nella figura 28 viene riportata nuovamente la distribuzione dei TPMs per ogni campione dopo la rimozione dei geni mitocondriali.

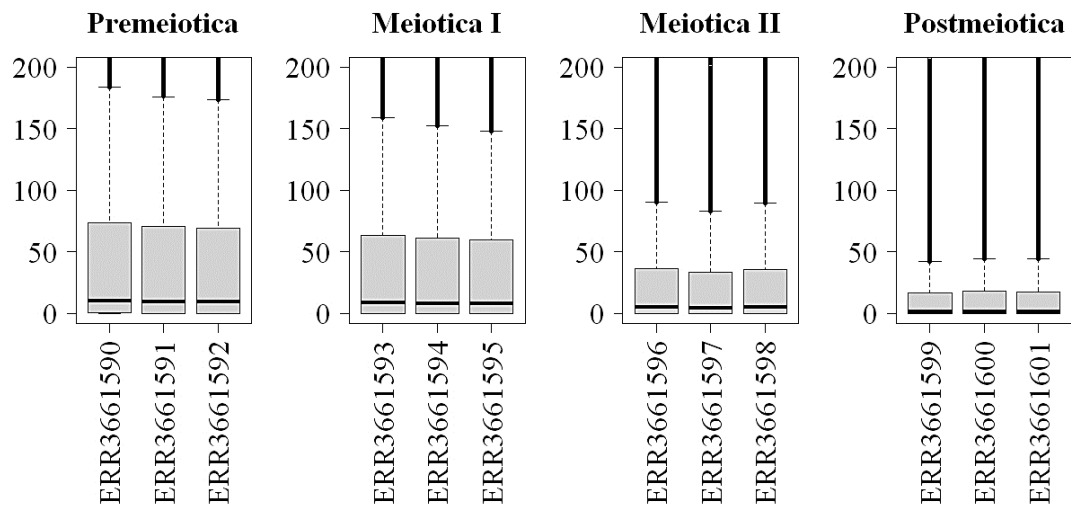


Figura 28: Distribuzione dei TPMs per ogni campione dopo la rimozione dei geni mitocondriali.

6.4 Espressione Genica prima della rimozione dei geni mitocondriali

L'espressione genica dei geni coinvolti nel processo di determinazione sessuale di *A. gambiae* viene riportata in figura 29. I livelli di espressione per ogni campione vengono rappresentati mediante dei grafici a barre e considerando una scala logaritmica per consentire una distribuzione più uniforme nello spazio cartesiano. La trasformazione in scala logaritmica non alcun effetto sull'analisi ed è utilizzata solo a scopo grafico.

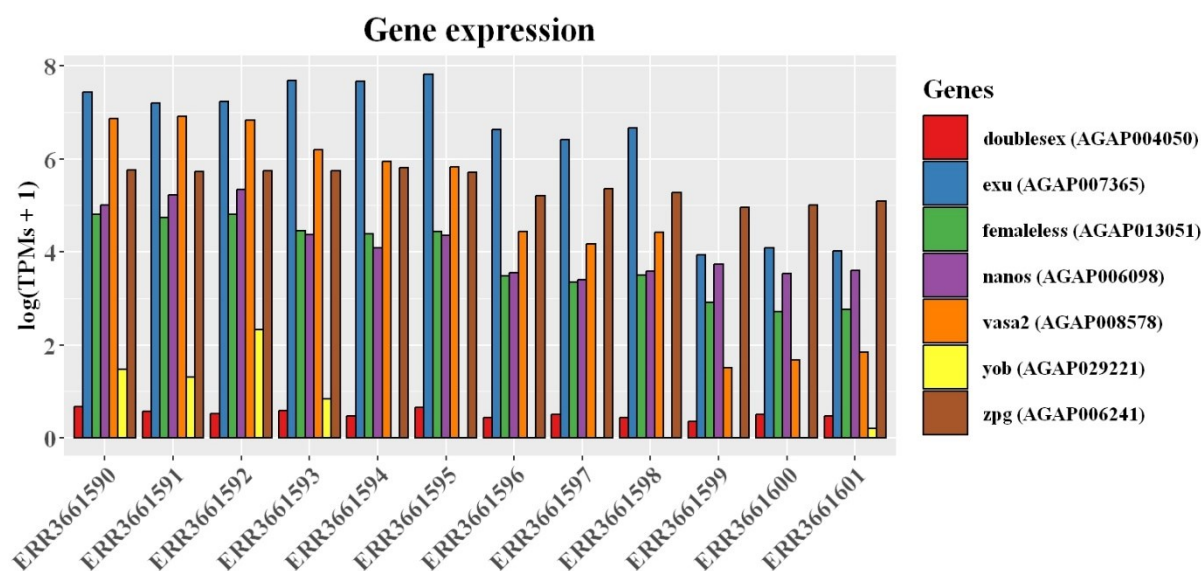


Figura 29: Rappresentazione dell'espressione genica.

I valori dei TPMs per ogni campione, rappresentati in figura 29, vengono riportati nelle tabelle successive. La tabella 8 riporta i valori dei geni analizzati *dsx*, *fle* e *yob*, mentre la tabella 9 riporta i valori dei promotori *exu*, *nanos*, *vasa2* e *zpg*.

Samples	<i>dsx</i>	<i>fle</i>	<i>yob</i>
ERR3661590	0,944061	121,6193	3,356256
ERR3661591	0,782719	113,3103	2,723773
ERR3661592	0,670879	120,5054	9,302406
ERR3661593	0,810316	84,99775	1,295349
ERR3661594	0,612546	79,77238	0
ERR3661595	0,909366	83,37701	0
ERR3661596	0,535654	31,61506	0
ERR3661597	0,649628	27,34243	0
ERR3661598	0,557871	31,90366	0
ERR3661599	0,42539	17,26803	0
ERR3661600	0,654711	14,08187	0
ERR3661601	0,598529	14,89928	0,224303

Tabella 8: TPMs dei geni *dsx*, *fle* e *yob* in ogni campione.

Samples	<i>exu</i>	<i>nanos</i>	<i>vasa2</i>	<i>zpg</i>
ERR3661590	1688,274	147,3297	954,7101	317,1092
ERR3661591	1326,213	182,8668	1005,089	307,2484
ERR3661592	1374,759	207,0257	926,6482	308,2708
ERR3661593	2152,921	78,25807	491,2718	310,7811
ERR3661594	2115,921	58,14547	381,861	331,9531
ERR3661595	2490,904	77,20158	336,757	298,7586
ERR3661596	759,0795	34,05988	83,72922	180,3569
ERR3661597	606,4241	28,95208	63,43159	211,2839
ERR3661598	780,9656	34,68768	82,57686	192,7336
ERR3661599	50,2572	40,7163	3,508844	140,7671
ERR3661600	58,37056	33,33166	4,373252	147,3193
ERR3661601	54,13613	35,70312	5,281428	162,1383

Tabella 9: TPMs dei promotori *exu*, *nanos*, *vasa2* e *zpg* in ogni campione.

6.5 Espressione genica dopo la rimozione dei geni mitocondriali

Nella figura 30 vengono riportati i livelli di espressione degli stessi geni e promotori analizzati nella sezione 6.4, in questo caso dopo aver rimosso il genoma mitocondriale.

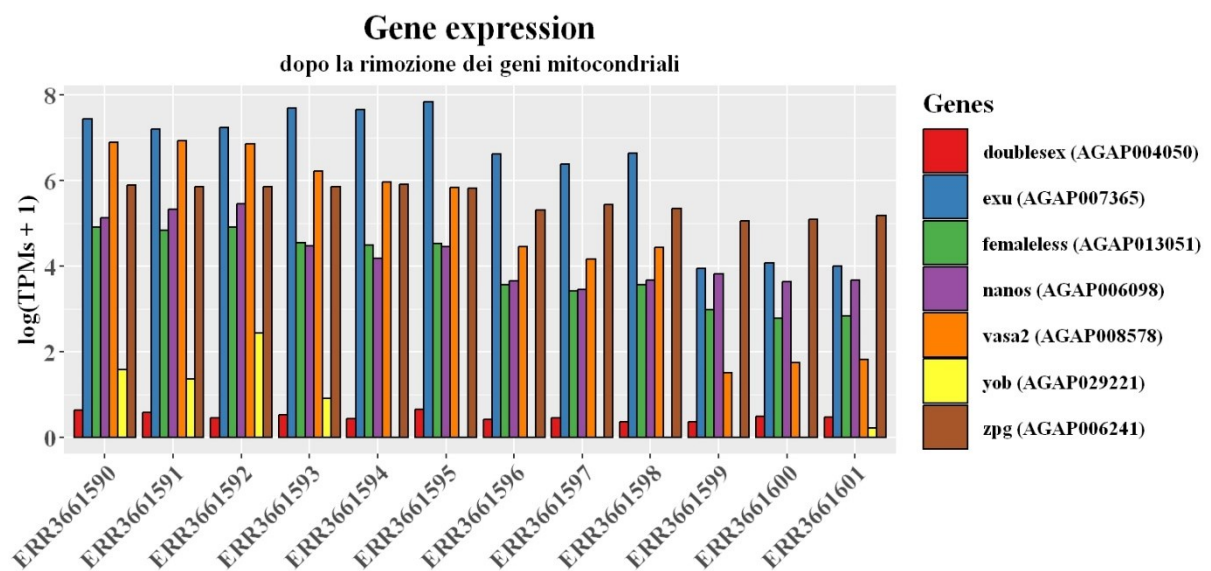


Figura 30: Rappresentazione dell'espressione genica dopo la rimozione dei geni mitocondriali.

Nelle tabelle 10 e 11 vengono riportati i valori dei TPMs dei geni e promotori plottati in figura 30.

Samples	<i>dsx</i>	<i>fle</i>	<i>yob</i>
ERR3661590	0,883764	134,1462	3,811476
ERR3661591	0,776925	125,2744	2,890404
ERR3661592	0,579996	133,6826	10,45489
ERR3661593	0,700091	93,32599	1,461741
ERR3661594	0,54427	87,38725	0
ERR3661595	0,930219	91,59466	0
ERR3661596	0,509278	34,29019	0
ERR3661597	0,568392	29,31793	0
ERR3661598	0,435525	34,03627	0
ERR3661599	0,433561	18,60546	0
ERR3661600	0,630422	15,13373	0
ERR3661601	0,606908	15,8575	0,246396

Tabella 10: TPMs dei geni *dsx*, *fle* e *yob* in ogni campione dopo la rimozione del genoma mitocondriale.

Samples	<i>exu</i>	<i>nanos</i>	<i>vasa2</i>	<i>zpg</i>
ERR3661590	1700,213	165,6603	975,271	356,8701
ERR3661591	1328,698	204,3709	1022,392	345,6778
ERR3661592	1372,093	230,9105	939,049	345,2397
ERR3661593	2160,364	86,89847	498,1004	346,5113
ERR3661594	2094,153	64,13721	387,4761	365,2255
ERR3661595	2503,985	85,18676	337,6781	331,0819
ERR3661596	748,3727	37,61226	85,0451	199,0632
ERR3661597	591,7509	30,8126	63,04456	229,3927
ERR3661598	759,8417	37,91249	82,67832	207,0195
ERR3661599	50,50001	44,49875	3,513864	154,2494
ERR3661600	57,2232	36,48449	4,716332	159,825
ERR3661601	53,36208	38,57898	5,171028	175,2969

Tabella 11: TPMs dei promotori *exu*, *nanos*, *vasa2* e *zpg* in ogni campione dopo la rimozione del genoma mitocondriale.

6.6 Espressione Genica differenziale

Le tre dispersioni previste dall'utilizzo del GLM sono rappresentate in figura 31.

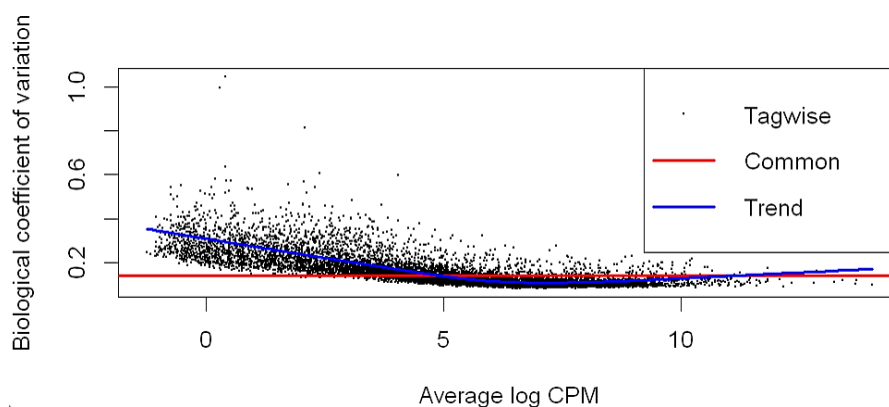


Figura 31: Rappresentazione delle tre dispersioni calcolate in edgeR.

I risultati dell'analisi dell'espressione genica differenziale eseguita con edgeR sono riportati nella tabella 12.

Totale di geni: 13.845	Up	Down
Meiotica I vs Premeiotica (2 vs 1)	1.763	1.373
Meiotica II vs Premeiotica (3 vs 1)	2.793	3.132
Postmeiotica vs Premeiotica (4 vs 1)	2.917	3.750
Meiotica I vs Meiotica II (3 vs 2)	1.730	2.173
Meiotica I vs Postmeiotica (4 vs 2)	2.360	3.323
Meiotica II vs Postmeiotica (3 vs 4)	813	2.110

Tabella 12: Numeri di geni differenzialmente espressi, partendo da un totale di 13.845.

Dai dati della tabella 12 si osserva come si è partito da un numero totale di geni pari a 13.845 e quelli differenzialmente espressi sono stati rappresentati mediante dei volcano plot nelle figure 32, 33 e 34. Per ogni confronto a coppie sono state riportate le soglie scelte per il FC e per il FDR e sono stati indicati in verde i geni sovra-espressi e in rosso quelli sotto-espressi.

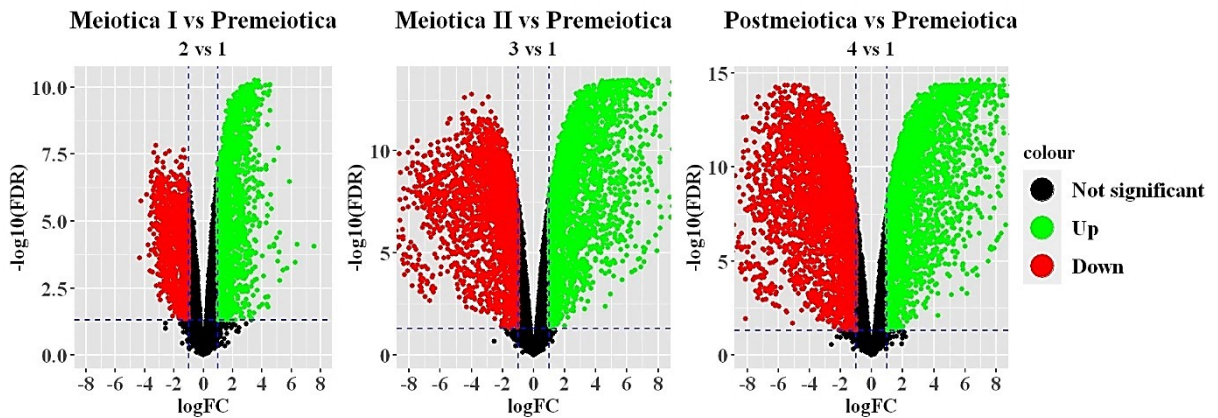


Figura 32: Volcano plot dei confronti a coppie effettuati per valutare i geni up-regulated e down-regulated rispetto alla popolazione premeiotica.

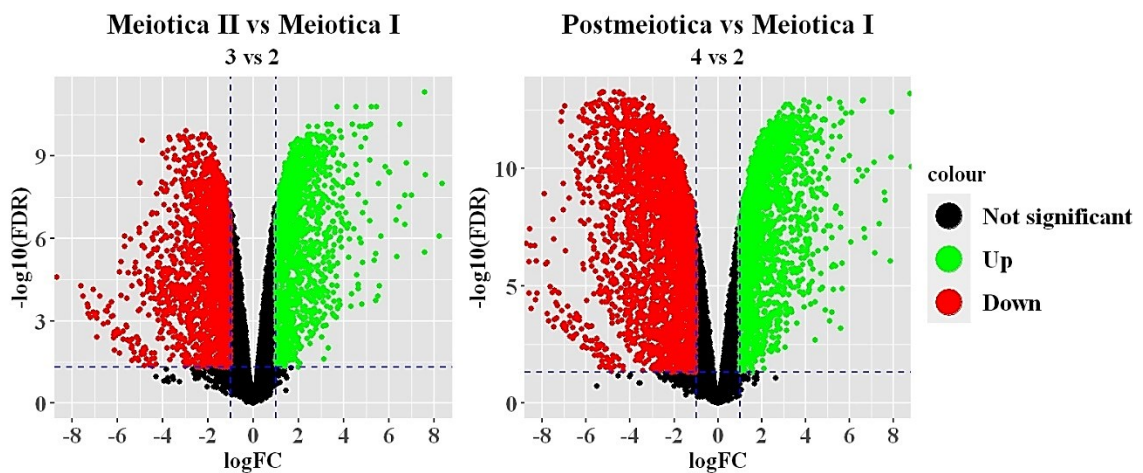


Figura 33: Volcano plot dei confronti a coppie effettuati per valutare i geni up-regulated e down-regulated rispetto alla popolazione meiotica I.

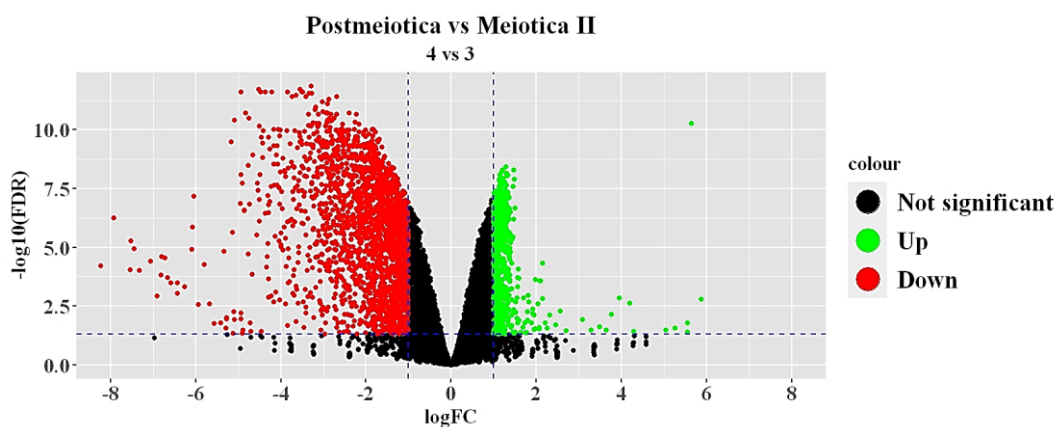


Figura 34: Volcano plot dei confronti a coppie effettuati per valutare i geni up-regulated e down-regulated rispetto alla popolazione meiotica II.

6.7 Arricchimento funzionale

I risultati dell'analisi di arricchimento funzionale vengono riportati per ogni confronto a coppie effettuato, distinguendo tra annotazione dei geni up-regulated e down-regulated. Poiché Gene Ontology permette di classificare i prodotti genici in tre ambiti fondamentali: molecular function, biological process e cellular component vengono riportati i risultati di tutte e tre le ontologie. I risultati dell'analisi GO sono mostrati nelle figure 35,36,37,38,39 e 40.

La figura 35 riporta i risultati del confronto meiotica I vs premeiotica.

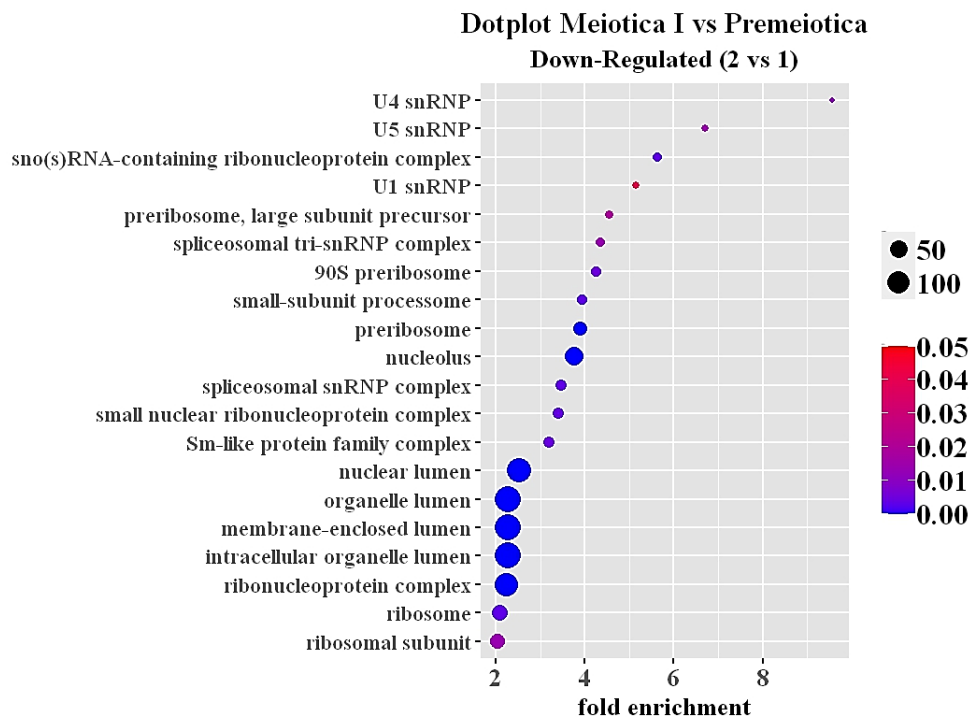
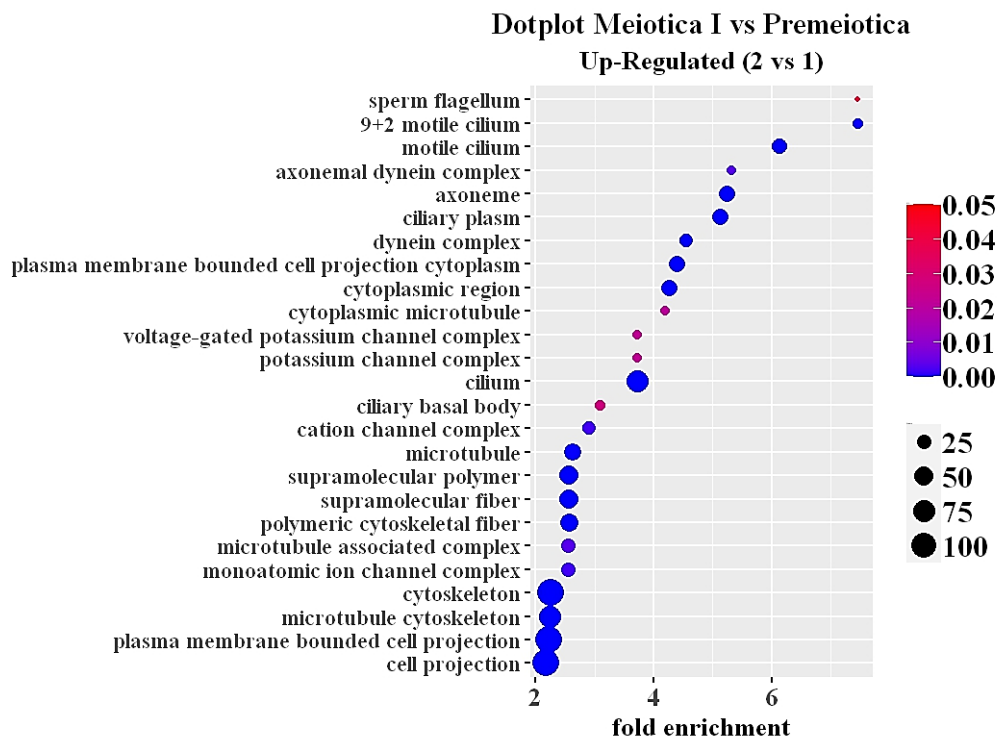


Figura 35: Annotazione dei geni del confronto Meiotica I vs Premeiotica. In alto si ha l'annotazione dei geni up-regulated, in basso quella dei geni down-regulated.

La figura 36 riporta i risultati del confronto meiotica II vs premeiotica.

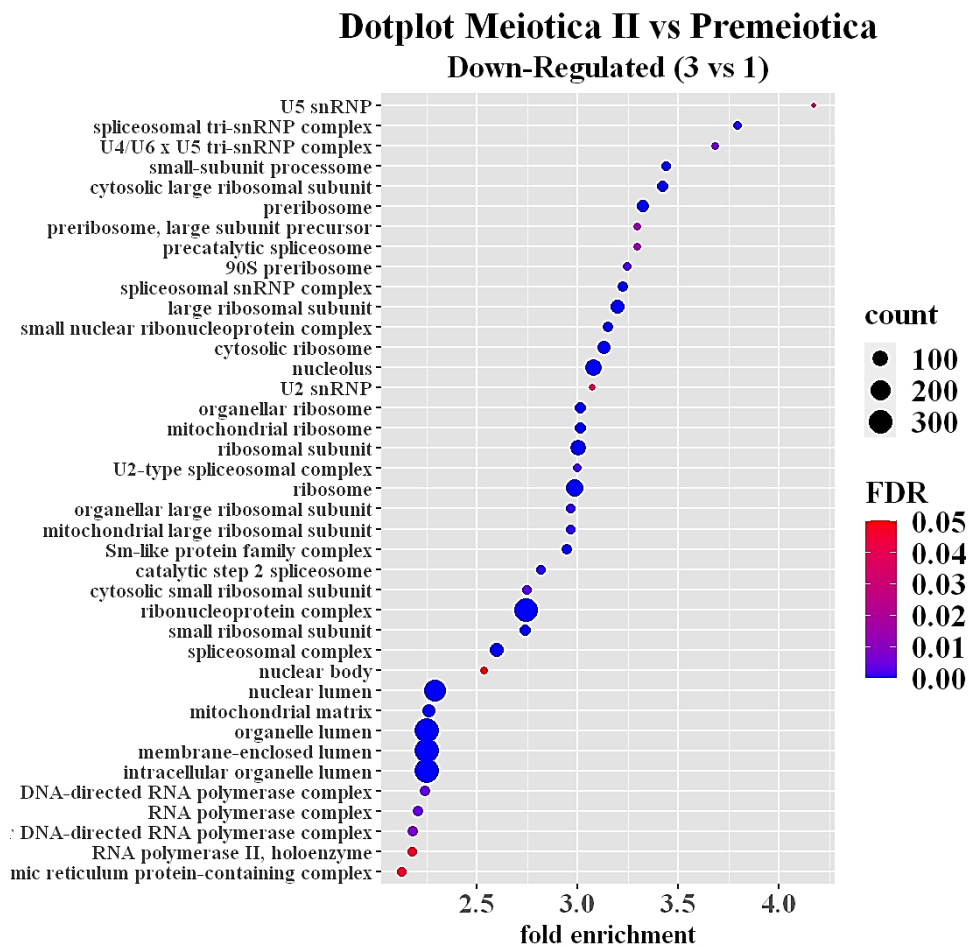
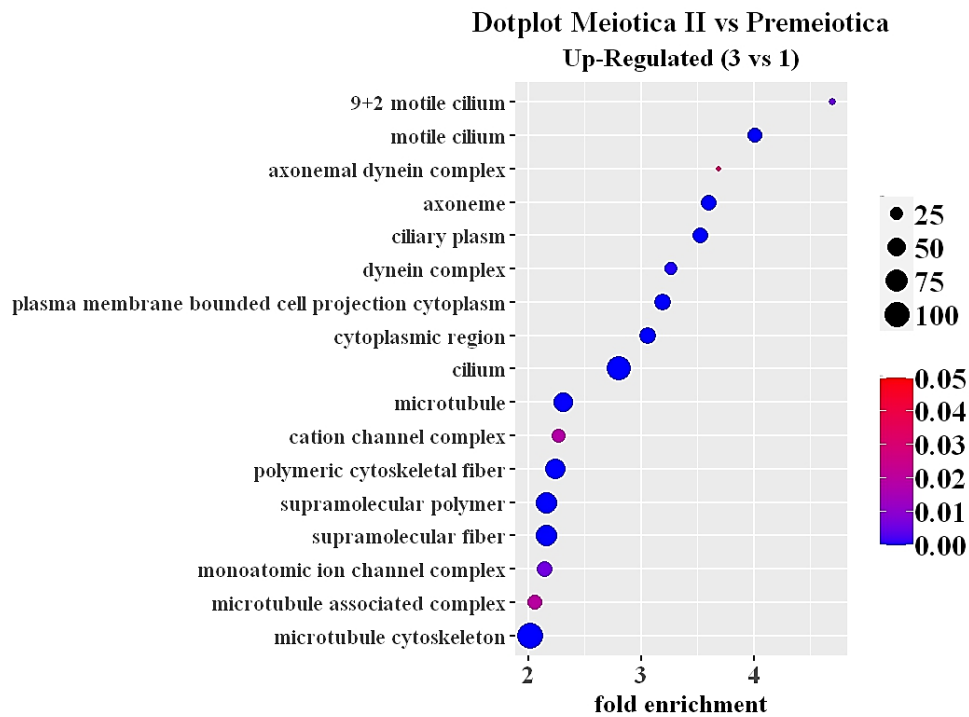


Figura 36: Annotazione dei geni del confronto Meiotica II vs Premeiotica. In alto si ha l'annotazione dei geni up-regulated, in basso quella dei geni down-regulated.

La figura 37 riporta i risultati del confronto postmeiotica vs premeiotica.

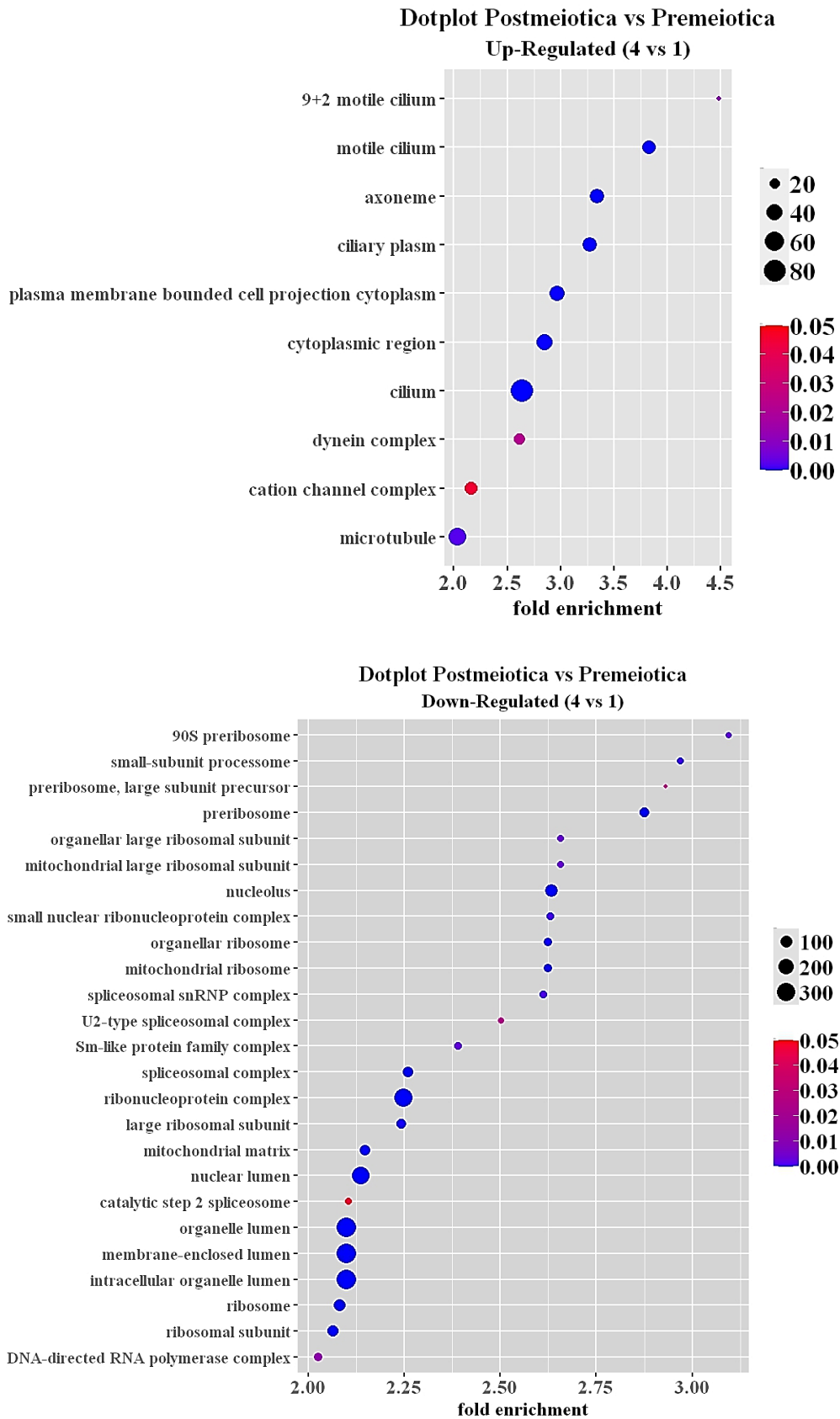


Figura 37: Annotazione dei geni del confronto Postmeiotica vs Premeiotica. In alto si ha l'annotazione dei geni up-regulated, in basso quella dei geni down-regulated.

La figura 38 riporta i risultati del confronto meiotica II vs meiotica I.

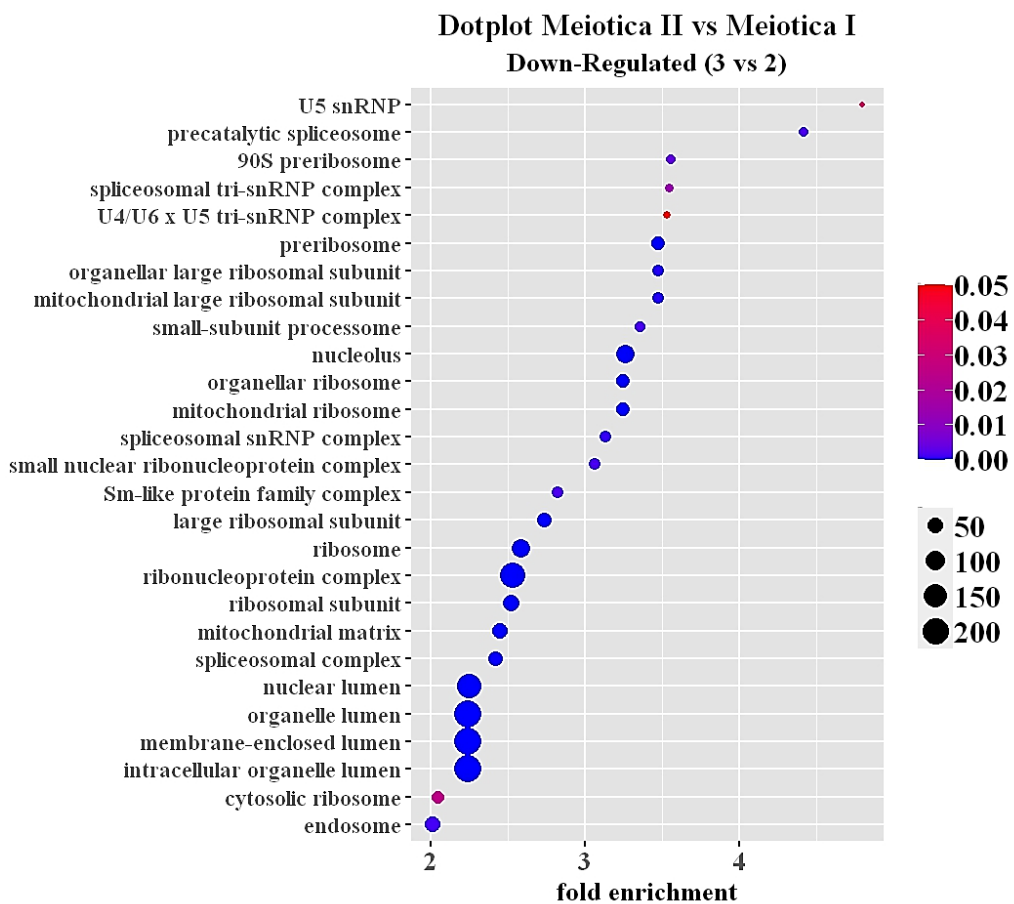
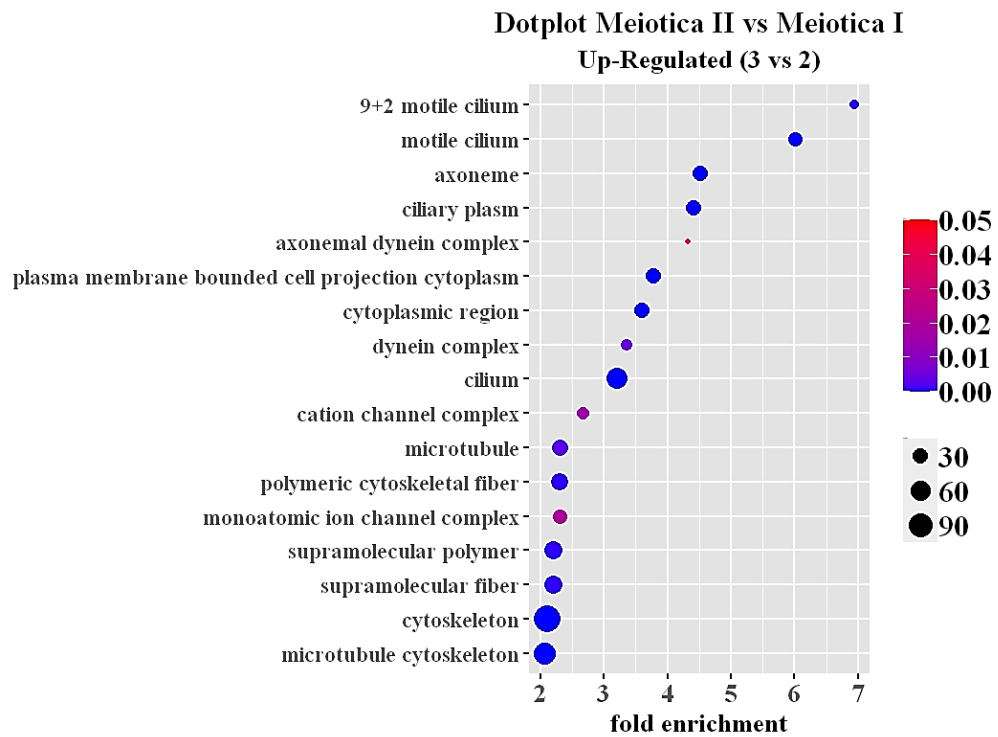


Figura 38: Annotazione dei geni del confronto Meiotica II vs Meiotica I. In alto si ha l'annotazione dei geni up-regulated, in basso quella dei geni down-regulated.

La figura 39 riporta i risultati del confronto postmeiotica vs meiotica I.

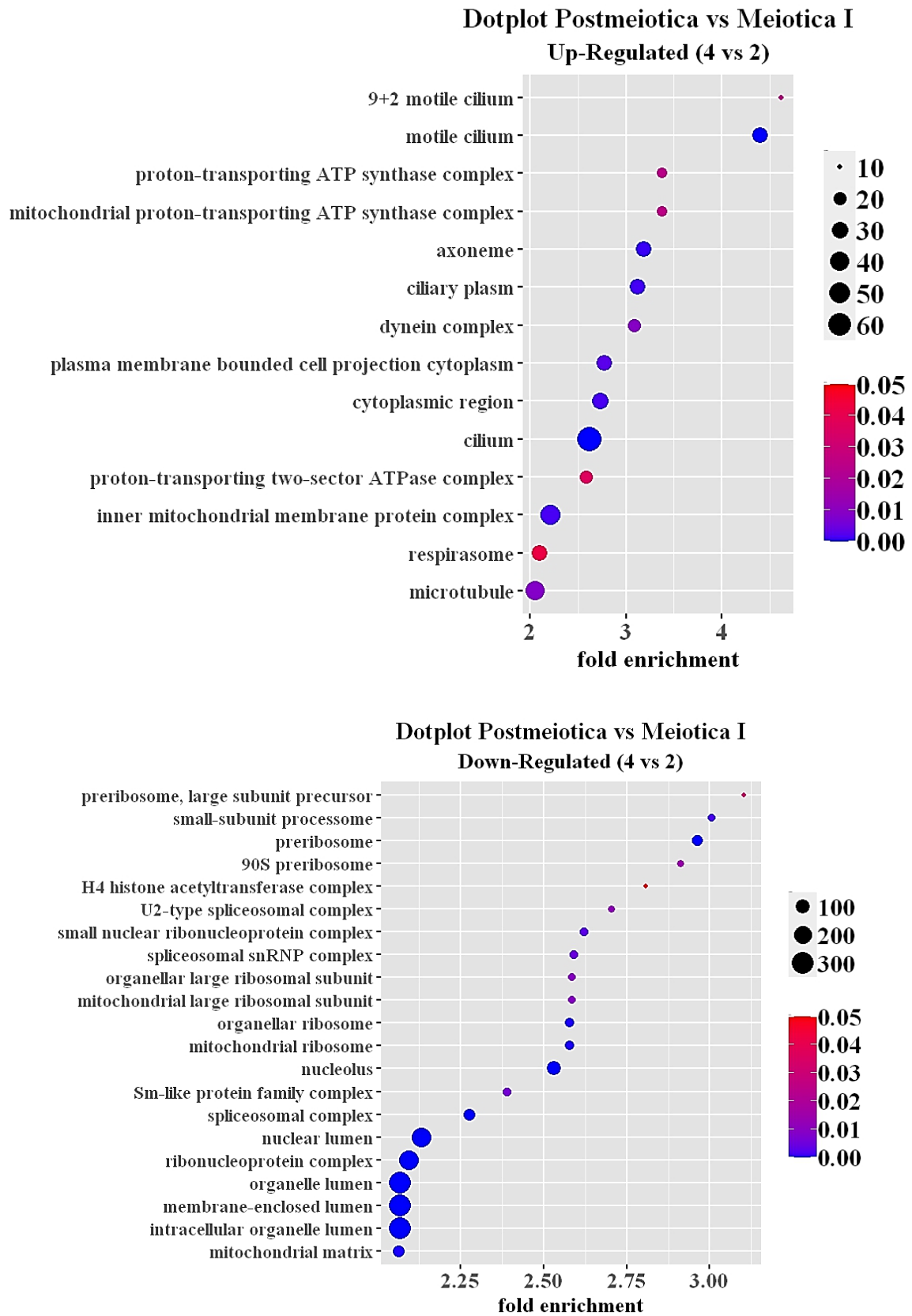


Figura 39: Annotazione dei geni del confronto Postmeiotica vs Meiotica I. In alto si ha l'annotazione dei geni up-regulated, in basso quella dei geni down-regulated.

La figura 40 riporta i risultati del confronto postmeiotica vs meiotica II.

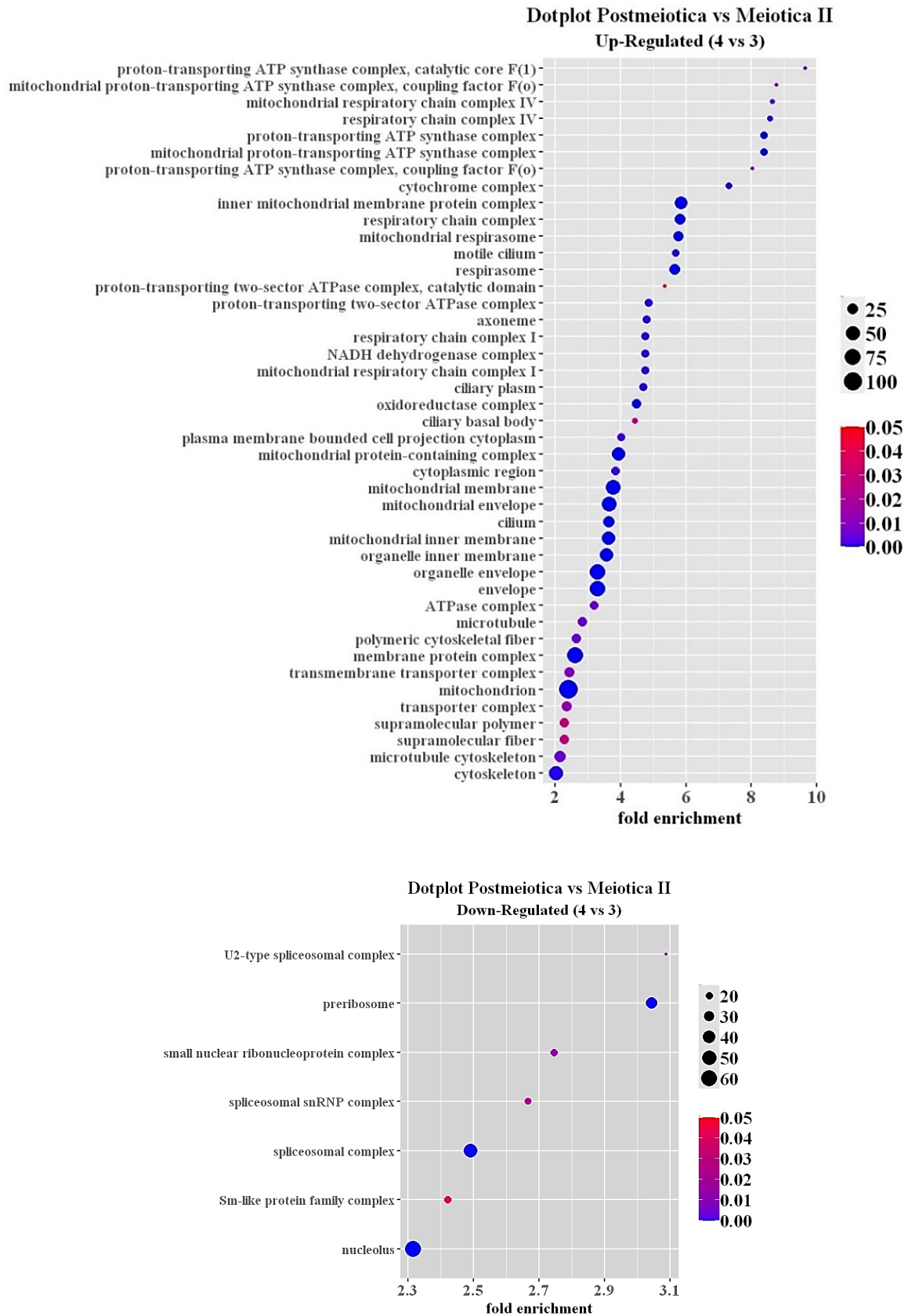


Figura 40: Annotazione dei geni del confronto Postmeiotica vs Meiotica II. In alto si ha l'annotazione dei geni up-regulated, in basso quella dei geni down-regulated.

Capitolo 7

Discussione

Il sequenziamento dell'RNA consente di analizzare e quantificare il trascrittoma cellulare, ovvero l'insieme delle molecole di RNA presenti in una cellula; in particolare, l'RNA messaggero (mRNA) contiene l'informazione necessaria per codificare le proteine. Lo studio dei trascritti di RNA consente la caratterizzazione di una cellula in una determinata condizione, definendo quali geni sono espressi, qual è il loro livello di trascrizione e in quale momento vengono attivati.

7.1 Quality control

Il controllo di qualità viene effettuato su files in formato FASTQ, ognuno dei quali è composto da 4 righe:

- una riga che identifica la sequenza, il cui nome è preceduto da @;
- la sequenza;
- un separatore che è semplicemente un segno + seguito dal nome della sequenza;
- punteggi di qualità rappresentati in caratteri ASCII.

Un esempio di FASTQ file è riportato in figura 41. Il file preso in considerazione è relativo alla forward read del primo campione ERR3661590.

```
@ERR3661590.1 1 length=100  
AGCGTAAACAAAACAGAGCAGCCCACTCGCGCACCTCCATCATCTACCTCATCTTTCGCAGCAAATTTAGCTTCTCCTCGTGATTTTCGATCATT  
+ERR3661590.1 1 length=100  
<BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB/ <<BFFB<FFBBBBBBBBB/FFBBBBBBBBBBBBBF/FBFFBFF
```

Figura 41: FASTQ file del primo campione ERR3661590.

Di tutte le statistiche generate dal tool FASTQC sono state analizzate nel dettaglio *per sequence quality score* e *sequence length distribution*. Il modulo *per sequence quality score*, nei report HTML, è rappresentato su un grafico bidimensionale in cui sull'asse x sono riportati i valori medi di qualità e sull'asse y il numero di sequenze con tale media. Questi dati consentono di osservare se una porzione significativa di reads è di bassa qualità.

Poiché il punteggio di qualità è calcolato come:

$$Q = -10 \log_{10}(p)$$

dove p indica la probabilità che la base chiamata sia errata, punteggi Q alti indicano una minore probabilità di errore. Punteggi $Q30$, ad esempio, prevedono che la probabilità di chiamata errata di una base sia di 1 su 1000 e che l'accuratezza sia del 99.9%. Se la qualità media più frequentemente osservata è al di sotto di 27, è generato un segnale di "warning" ed equivale ad un tasso di errore dello 0.2%, mentre se la qualità media osservata è inferiore a 20, allora il tasso d'errore è dell'1% e ciò indica una bassa qualità dei dati.

Nella figura 21 sono stati riportati i valori medi dei punteggi di qualità in maniera tale da confrontarli simultaneamente in tutti i campioni disponibili. La qualità media è stata calcolata distinguendo tra forward e reverse reads dello stesso campione. Nella figura 21a sono riportati i Phred scores. Tutti i punteggi di qualità sono compresi tra 35 e 36. Nella figura 21b vengono riportati i dati dopo aver effettuato l'operazione di trimming. In questo caso si osserva come i Phred scores siano quasi completamente identici, con un valore medio di circa 36 in tutti i campioni e una sovrapposizione dei punti rispetto alla figura 21a. Nel caso in esame, sebbene i punteggi siano già superiore a 30 prima del processamento con Trimmomatic, l'operazione di trimming non fa altro che migliorare ancor di più la qualità dei dati uniformando gli scores in tutti i campioni.

L'altra statistica analizzata è la *sequence length distribution* in cui viene riportata la distribuzione delle dimensioni dei frammenti nel campione analizzato. I report HTML raffigurano un grafico bidimensionale in cui sugli assi x e y vengono riportati rispettivamente la lunghezza della sequenza e le quantità relative a ciascuna lunghezza. È generato un messaggio di "warning" se le sequenze non sono della stessa lunghezza o un messaggio di "fail" se una delle sequenze ha lunghezza zero.

Nella figura 22 vengono rappresentati i valori medi delle lunghezze delle sequenze distinguendo sempre tra forward reads e reverse. Nella figura 22a si hanno le lunghezze prima del trimming e il valore medio è compreso tra 98 e 100 in tutti i campioni. Nella figura 22b le lunghezze medie riportate sono ottenute dopo la rimozione degli adattatori e si osserva, in questo caso, una minore differenza di lunghezza in quasi tutti i campioni, che invece era presente nella figura precedente, ad eccezione del campione ERR3661601 le cui lunghezze sono sovrapposte.

Data la buona qualità dei dati di partenza è stato possibile applicare parametri stringenti per il processamento dei dati con Trimmomatic, senza che ciò portasse a scartare un numero elevato di reads.

7.2 Allineamento al genoma di riferimento

L'utilizzo di Samtool dopo l'allineamento al genoma di riferimento ha permesso di valutare la quantità di duplicati e reads correttamente accoppiate. Il campo dei duplicati indica che la sequenza letta è stata identificata come non univoca, dunque, il file di allineamento contiene almeno un'altra read identica. La presenza di numerosi duplicati potrebbe essere associata anche all'amplificazione dello stesso frammento di cDNA nella PCR. L'eventuale rimozione dei duplicati, in questo caso, sarebbe necessaria per ridurre i biases associati al processo di amplificazione. Questo porterebbe ad avere un numero minore di reads da mappare e una riduzione del tempo di allineamento. La tabella 4 riporta il numero di duplicati sul totale per ogni campione. La figura 23 rappresenta gli stessi valori in percentuale ed è possibile osservare come in tutti i campioni ci sia un valore di duplicati minore del 25%, ad eccezione dei campioni della popolazione postmeiotica ERR3661599, ERR3661600 e ERR3661601.

La percentuale di reads correttamente accoppiate indica che entrambe le reads sono allineate e le coppie si trovano entro una certa distanza. Guardando i dati in tabella 5 e le rappresentazioni in figura 24 si osserva un'alta percentuale di reads correttamente accoppiate. La percentuale, infatti, è del 98% in tutti i campioni, ciò significa che è stato ottenuto un buon tasso di allineamento dal tool HISAT2.

Il conteggio dei duplicati e delle reads correttamente accoppiate viene riportato nuovamente dopo aver rimosso i geni mitocondriali grazie all'utilizzo di SAMtool. I dati della tabella 7 mostrano come il tasso di duplicati sia minore dopo il filtraggio in tutti i campioni, mentre la percentuale di reads correttamente accoppiate rimanga quasi invariata. La diminuzione dei duplicati dimostra come lo step di rimozione dei geni mitocondriali migliori la robustezza dei dati necessaria per l'esecuzione delle analisi a valle, nonostante il contributo sul totale delle reads sia basso in ogni campione come mostrato nella tabella 6.

7.3 Quantificazione dei trascritti

Negli esperimenti di RNA-seq il numero di reads allineate al genoma di riferimento viene utilizzato per quantificare i livelli di espressione dei vari geni. In questo studio l'aggregazione

delle reads mappate è stata effettuata con featureCounts che ha generato dei conteggi grezzi non direttamente paragonabili a causa della presenza di biases che si verificano tra i campioni di un esperimento definiti come “between-sample”. La corretta comparazione è ottenuta utilizzando come unità di espressione genica i Transcript per Million (TPM). I TPMs normalizzano prima per la lunghezza del gene e poi per la profondità di sequenziamento che può differire significativamente tra i campioni. Il controllo di un calcolo effettuato correttamente, può essere fatto constatando che la somma di tutti i TPMs in ogni campione sia uguale ad un milione.

La distribuzione dei TPMs è rappresentata in figura 26. Ci si aspetta che il numero di geni espressi debba essere approssimativamente uguale per i campioni appartenenti allo stesso gruppo. Dai boxplots in figura si osserva come effettivamente la distribuzione dei TPMs sia molto simile nei campioni appartenenti alla stessa popolazione e rimanga tale anche dopo la rimozione dei geni mitocondriali come riportato nella figura 28.

Nella figura 25 è rappresentato il risultato ottenuto eseguendo la PCA sui TPMs calcolati. L’analisi delle componenti principali aiuta a identificare i campioni per un controllo di qualità e permette di valutare le variazioni e le distanze esistenti. L’ampio set di conteggi per ogni gene è trasformato in un insieme più piccolo di componenti principali ortogonali. Nel caso della figura 25 è possibile notare come la varianza totale dei dati è spiegata principalmente dalle prime due componenti. La prima componente rappresentata sull’asse x spiega il 95.5% della varianza e la seconda componente sull’asse y, invece, spiega il 2.1%. Poiché la nostra matrice di dati riportava i geni sulle righe e i campioni delle quattro popolazioni sulle colonne, l’analisi è stata effettuata sulle popolazioni che compongono il set di dati osservando un raggruppamento dei campioni che appartengono alla stessa popolazione. Quindi si ha una distanza minima tra le repliche della stessa popolazione e una distanza più marcata tra le popolazioni in esame che vengono rappresentate mediante colori differenti. Il risultato ottenuto è paragonabile a quello ottenuto in [72] (Figura 42). Taxiarchi et al. utilizzano lo stesso set di dati ma eseguono la PCA sui conteggi grezzi.

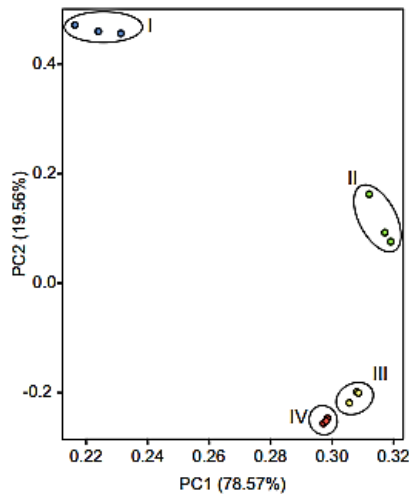


Figura 42: PCA eseguita da Taxiarchi et al. eseguita in [72] sullo stesso set di dati.

La PCA è rieseguita dopo la rimozione dei geni mitocondriali. I risultati vengono riportati nella figura 27. La percentuale di varianza spiegata dalla prima componente in questo caso è del 97%, mentre quella della seconda è dell' 1.4%. La rimozione dei geni mitocondriali non comporta una variazione nella distribuzione delle 4 popolazioni sul grafico e la relazione tra le repliche della stessa popolazione viene mantenuta.

7.4 Espressione genica prima della rimozione dei geni mitocondriali

La figura 29 riporta l'espressione genica dei principali geni coinvolti nel processo di determinazione sessuale in *A. gambiae* e dei geni i cui promotori sono stati usati nelle tecniche di gene drive. Per quanto riguarda i geni coinvolti nella determinazione del sesso:

- *doublesex* risulta poco espresso in tutti i campioni;
- *femaleless* è maggiormente espresso nella popolazione premeiotica, conserva una più bassa espressione nelle altre popolazioni ma il suo livello tende comunque a diminuire;
- *yob* ha un'espressione pari a 0 in tutti i campioni, ad eccezione della popolazione premeiotica, della prima replica della popolazione meiotica I e dell'ultima replica della postmeiotica. Nonostante sia espresso in queste repliche, il livello di espressione è basso.

Per quanto riguarda i geni i cui promotori sono usati nel gene drive si ha:

- *exu* è quello maggiormente espresso considerando tutti i geni presi in esame. La sua espressione è particolarmente alta nei campioni delle popolazioni premeiotica e

meiotica I e tende a diminuire nelle ultime due popolazioni, in particolare in quella postmeiotica;

- *nanos* ha una più alta espressione nella popolazione premeiotica e poi si osserva una riduzione dalla popolazione meiotica I fino ad arrivare ad un livello quasi costante nelle ultime due popolazioni.
- *vasa2* è particolarmente espresso nella popolazione premeiotica. La sua espressione però si riduce ad ogni popolazione successiva fino ad arrivare a valori piuttosto bassi nella popolazione postmeiotica;
- *zpg* ha un'espressione che non varia molto nelle popolazioni premeiotica e meiotica I, si riduce nella popolazione meiotica I e da qui rimane più o meno costante fino alla popolazione postmeiotica.

Un aspetto rilevante è la bassa espressione di *yob* in tutti i campioni, addirittura pari a 0, quindi, non espresso, in alcune repliche. Questo risulta essere particolarmente importante considerando che si tratta di un gene che si trova sul cromosoma Y e nel nostro caso l'RNA è stato estratto da *A. gambiae* di sesso maschile. La bassa espressione di *yob* potrebbe essere causata dalla MSC1. Come suggerito in [75], sarebbero necessari ulteriori studi sull'espressione dei geni legati al cromosoma Y durante la spermatogenesi e, inoltre, sarebbero da indagare maggiormente l'entità, la portata e i tempi esatti del processo. Una down-regulation dei geni Y-linked potrebbe essere dovuta alla formazione di cromatina condensata che sembra inibire l'attività trascrizionale di geni legati al sesso durante gli stadi meiotici della spermatogenesi [76].

Guardando i dati risulta espresso anche *fle*, che sappiamo avere un ruolo nel percorso di determinazione del sesso e regolare lo splicing di *dsx* nelle femmine. Nei maschi, però, *fle* pur essendo espresso è inattivo e questo è stato dimostrato in [57] in cui l'abbattimento di *fle* ha provocato una letalità specifica per le femmine negli embrioni di *A. gambiae*, ma apparentemente nessun effetto sullo sviluppo dei maschi. L'andamento di *dsx* non segue quello di *fle*, nonostante quest'ultimo ne controlli lo splicing; questo può essere dovuto alla possibilità che il gene *fle* richieda la presenza di un cofattore. Non è ancora chiaro se la presenza di un cofattore sia necessaria per lo splicing di *dsx*, così come avviene in *D. melanogaster*, oppure se è ipotizzabile l'idea che il solo legame di *fle* sia sufficiente.

Zpg è espresso, come ci aspettiamo, ed è infatti il promotore più utilizzato per esprimere il gene drive. Un confronto fra i tassi di fertilità e homing in individui eterozigoti con un sistema gene drive basato su CRISPR guidato da *vasa2* e *zpg*, nello stesso identico locus, ha mostrato una migliore fertilità nei costrutti guidati da *zpg* [53]. Questa differenza è stata osservata anche in [77], in cui sono stati testati altri costrutti basati sui promotori *exu* e *nanos*, i quali hanno

determinato una riduzione della fecondità anche nei maschi rispetto all'utilizzo di *zpg*. Inoltre, un problema principale associato al promotore *nanos* è che provoca dei problemi di resistenza insieme a *vasa2* ed *exu*.

7.5 Espressione genica dopo la rimozione dei geni mitocondriali

A seguito della rimozione dei geni mitocondriali sono stati ricalcolati i valori dei livelli di espressione dei geni *doublesex*, *femaleless* e *yob* e dei promotori *exu*, *nanos*, *vasa2* e *zpg*. La figura 30 e i dati riportati nelle tabelle 10 e 11 mostrano come non ci sia, in nessun caso, una variazione nel pattern di espressione di questi geni dopo la rimozione del genoma mitocondriale, questo è principalmente dovuto al basso contributo del genoma mitocondriale sul numero totale di reads. Le piccole variazioni che si riscontrano non sono significative al punto tale da ricavare informazioni differenti da quelle apprese dall'analisi effettuata prima della rimozione dei geni mitocondriali.

7.6 Espressione genica differenziale

L'attivazione o inattivazione dell'espressione genica è influenzata dal differenziamento cellulare durante lo sviluppo, dalla regolazione del ciclo cellulare e dalla risposta a mediatori esterni. L'espressione di un gene determina un fenotipo morfologico e funzionale. Finora il focus dell'analisi è stato la valutazione dell'attivazione o inattivazione di geni in un tessuto specifico, in particolari momenti dello sviluppo e lo studio di quanto il singolo gene era attivo. Nel caso dell'analisi differenziale, invece, si valuta l'espressione genica confrontando i profili trascrizionali di due o più tessuti, o come nel nostro caso, profili trascrizionali dello stesso tipo cellulare ma in condizioni diverse, ad esempio per diverse fasi dello sviluppo.

L'analisi dell'espressione genica differenziale è stata condotta utilizzando le counts prodotte da featureCounts, dopo aver rimosso il genoma mitocondriale.

7.7 Arricchimento funzionale

L'analisi mostra una presenza significativa di termini associati al citoscheletro, ai microtubuli, alle ciglia, alle dineine e altre componenti cellulari. Si tratta di termini coinvolti principalmente nel mantenimento della struttura e nell'attività/motilità degli spermatozoi e presenti in tutte e quattro le popolazioni, principalmente in quelle meiotiche. Seguono numerosi termini legati a

componenti mitocondriali, alla loro attività e a complessi proteici transmembrana essenziali per la respirazione cellulare, maggiormente presenti nella popolazione postmeiotica.

Molti termini si riferiscono alla formazione di ATP, una fonte di energia fornita dai mitocondri per consentire alla cellula lo svolgimento delle funzioni vitali; in questo caso è possibile che la richiesta di energia sia associata alla motilità delle cellule spermatiche. La presenza maggiore di questi termini nella popolazione postmeiotica è coerente con la presenza dei componenti mitocondriali nella stessa popolazione.

Altri termini sono associati alla formazione dei ribosomi necessari per la sintesi proteica e alla presenza di geni legati a complessi ribonucleoproteici (RNA e proteine leganti l'RNA) coinvolti nell'elaborazione dei trascritti di RNA e nella formazione dello spliceosoma che catalizza lo splicing degli esoni codificanti per produrre mRNA maturo. La presenza di questi termini principalmente nelle popolazioni premeiotiche e meiotiche potrebbe essere ricollegata alla MSC1. Poiché la MSC1 porta ad una riduzione/inibizione dell'attività trascrizionale, è possibile ipotizzare una conseguente riduzione del contenuto di RNA e quindi anche di complessi ribonucleoproteici nelle popolazioni postmeiotiche.

Bibliografia

- [1] «World malaria report 2023». Consultato: 12 febbraio 2024. [Online]. Disponibile su: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023>
- [2] «Global technical strategy for malaria 2016–2030». Consultato: 28 novembre 2023. [Online]. Disponibile su: <https://www.who.int/publications-detail-redirect/9789241564991>
- [3] «World malaria report 2022». Consultato: 28 novembre 2023. [Online]. Disponibile su: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2022>
- [4] A. Rossati, O. Bargiacchi, V. Kroumova, M. Zaramella, A. Caputo, e P. L. Garavelli, «Climate, environment and transmission of malaria», *Infez Med*, vol. 24, fasc. 2, pp. 93–104, giu. 2016.
- [5] S. I. Hay, D. L. Smith, e R. W. Snow, «Measuring malaria endemicity from intense to interrupted transmission», *The Lancet Infectious Diseases*, vol. 8, fasc. 6, pp. 369–378, giu. 2008, doi: 10.1016/S1473-3099(08)70069-0.
- [6] M. della Salute, «Malaria». Consultato: 24 gennaio 2024. [Online]. Disponibile su: <https://www.salute.gov.it/portale/malattieInfettive/dettaglioSchedeMalattieInfettive.jsp?lingua=italiano&id=212&area=Malattie%20infettive&menu=indiceAZ&tab=1>
- [7] «African malaria mosquito - *Anopheles gambiae* Say». Consultato: 24 gennaio 2024. [Online]. Disponibile su: https://entnemdept.ufl.edu/creatures/aquatic/Anopheles_gambiae.htm
- [8] «Malaria - Infezioni», Manuale MSD, versione per i pazienti. Consultato: 24 gennaio 2024. [Online]. Disponibile su: <https://www.msdmanuals.com/it-it/casa/infezioni/infezioni-parassitarie-protozoi-extraintestinali/malaria>
- [9] «CDC - Parasites - Malaria». Consultato: 24 gennaio 2024. [Online]. Disponibile su: <https://www.cdc.gov/parasites/malaria/index.html>
- [10] C.-C. for D. C. and Prevention, «CDC - Malaria - About Malaria - Biology». Consultato: 24 gennaio 2024. [Online]. Disponibile su: <https://www.cdc.gov/malaria/about/biology/index.html>
- [11] «Malaria mosquitoes (*Anopheles*)», Biogents AG. Consultato: 24 gennaio 2024. [Online]. Disponibile su: <https://eu.biogents.com/anopheles-malaria-mosquitoes/>
- [12] C.-C. for D. C. and Prevention, «CDC - Malaria - About Malaria». Consultato: 24 gennaio 2024. [Online]. Disponibile su: https://www.cdc.gov/malaria/about/preventing_malaria.html

- [13] C. N. Ngonghala, «Assessing the impact of insecticide-treated nets in the face of insecticide resistance on malaria control», *Journal of Theoretical Biology*, vol. 555, p. 111281, dic. 2022, doi: 10.1016/j.jtbi.2022.111281.
- [14] J. P. Gutiérrez-Jara, K. Vogt-Geisse, e M. Cabrera, «Collateral Effects of Insecticide-Treated Nets on Human and Environmental Safety in an Epidemiological Model for Malaria with Human Risk Perception», *IJERPH*, vol. 19, fasc. 23, p. 16327, dic. 2022, doi: 10.3390/ijerph192316327.
- [15] T. H. Barker, J. C. Stone, S. Hasanoff, C. Price, A. Kabaghe, e Z. Munn, «Effectiveness of dual active ingredient insecticide-treated nets in preventing malaria: A systematic review and meta-analysis», *PLoS ONE*, vol. 18, fasc. 8, p. e0289469, ago. 2023, doi: 10.1371/journal.pone.0289469.
- [16] L. Mugenyi *et al.*, «Estimating the optimal interval between rounds of indoor residual spraying of insecticide using malaria incidence data from cohort studies», *PLoS One*, vol. 15, fasc. 10, p. e0241033, 2020, doi: 10.1371/journal.pone.0241033.
- [17] G. Fuseini *et al.*, «Improving the performance of spray operators through monitoring and evaluation of insecticide concentrations of pirimiphos-methyl during indoor residual spraying for malaria control on Bioko Island», *Malar J*, vol. 19, fasc. 1, p. 35, dic. 2020, doi: 10.1186/s12936-020-3118-y.
- [18] J. Yukich *et al.*, «Incremental cost and cost-effectiveness of the addition of indoor residual spraying with pirimiphos-methyl in sub-Saharan Africa versus standard malaria control: results of data collection and analysis in the Next Generation Indoor Residual Sprays (NgenIRS) project, an economic-evaluation», *Malar J*, vol. 21, fasc. 1, p. 185, dic. 2022, doi: 10.1186/s12936-022-04160-3.
- [19] C.-C. for D. C. and Prevention, «CDC - Malaria - Malaria Worldwide - How Can Malaria Cases and Deaths Be Reduced? - Larval Control and Other Vector Control Interventions». Consultato: 27 gennaio 2024. [Online]. Disponibile su: https://www.cdc.gov/malaria/malaria_worldwide/reduction/vector_control.html
- [20] S. H. P. P. Karunaratne e S. N. Surendran, «Mosquito control: A review on the past, present and future strategies», *J. Natn. Sci. Foundation Sri Lanka*, vol. 50, fasc. 0, p. 277, nov. 2022, doi: 10.4038/jnsfsr.v50i0.11244.
- [21] O. J. Adelaja, A. O. Oduola, A. T. Ande, O. O. Abiodun, e A. R. Adelaja, «Toxicity of insecticidal plant oils on the larval and adult stages of a major malaria vector (*Anopheles gambiae* Giles 1920)», *Parasitol Res*, vol. 122, fasc. 5, pp. 1071–1078, mag. 2023, doi: 10.1007/s00436-023-07806-6.

- [22] C. Kiuru *et al.*, «Effects of larval exposure to sublethal doses of ivermectin on adult fitness and susceptibility to ivermectin in *Anopheles gambiae* s.s.», *Parasites Vectors*, vol. 16, fasc. 1, p. 293, ago. 2023, doi: 10.1186/s13071-023-05888-w.
- [23] P. Jagannathan e A. Kakuru, «Malaria in 2022: Increasing challenges, cautious optimism», *Nat Commun*, vol. 13, fasc. 1, p. 2678, mag. 2022, doi: 10.1038/s41467-022-30133-w.
- [24] V. Jishtu e A. Kumar, «Chemoprophylaxis of Malaria in Special Population Groups and Challenges: What Have We Achieved So Far and the Way Forward», *Himalayan Journal of Medicine and Surgery*, vol. 3, fasc. 1, 2022.
- [25] S. S. Nundu *et al.*, «It Is Time to Strengthen the Malaria Control Policy of the Democratic Republic of Congo and Include Schools and School-Age Children in Malaria Control Measures», *Pathogens*, vol. 11, fasc. 7, p. 729, giu. 2022, doi: 10.3390/pathogens11070729.
- [26] M. González-Sanz, P. Berzosa, e F. F. Norman, «Updates on Malaria Epidemiology and Prevention Strategies», *Curr Infect Dis Rep*, vol. 25, fasc. 7, pp. 131–139, lug. 2023, doi: 10.1007/s11908-023-00805-9.
- [27] EpiCentro, «Malaria». Consultato: 25 novembre 2023. [Online]. Disponibile su: <https://www.epicentro.iss.it/malaria/>
- [28] P. E. Duffy, «Current approaches to malaria vaccines», *Current Opinion in Microbiology*, vol. 70, p. 102227, dic. 2022, doi: 10.1016/j.mib.2022.102227.
- [29] P. E. Duffy e J. Patrick Gorres, «Malaria vaccines since 2000: progress, priorities, products», *npj Vaccines*, vol. 5, fasc. 1, p. 48, giu. 2020, doi: 10.1038/s41541-020-0196-3.
- [30] A. Nadeem e W. Bilal, «Acceptance, availability and feasibility of RTS, S/AS01 malaria vaccine: A review of literature», *Asian Pac J Trop Med*, vol. 16, fasc. 4, p. 162, 2023, doi: 10.4103/1995-7645.374354.
- [31] J. G. Beeson *et al.*, «The RTS,S malaria vaccine: Current impact and foundation for the future», *Sci. Transl. Med.*, vol. 14, fasc. 671, p. eabo6646, nov. 2022, doi: 10.1126/scitranslmed.abo6646.
- [32] M. E. Tsoumani, C. Voyiatzaki, e A. Efstathiou, «Malaria Vaccines: From the Past towards the mRNA Vaccine Era», *Vaccines*, vol. 11, fasc. 9, p. 1452, set. 2023, doi: 10.3390/vaccines11091452.
- [33] S. Sang *et al.*, «Safety and immunogenicity of varied doses of R21/Matrix-MTM vaccine at three years follow-up: A phase 1b age de-escalation, dose-escalation trial in adults, children, and infants in Kilifi-Kenya», *Wellcome Open Res*, vol. 8, p. 450, ott. 2023, doi: 10.12688/wellcomeopenres.19795.1.

- [34] E. Takashima, M. Tachibana, M. Morita, H. Nagaoka, B. N. Kanoi, e T. Tsuboi, «Identification of Novel Malaria Transmission-Blocking Vaccine Candidates», *Front. Cell. Infect. Microbiol.*, vol. 11, p. 805482, nov. 2021, doi: 10.3389/fcimb.2021.805482.
- [35] «WHO malaria terminology, 2021 update», 2021.
- [36] K. P. Paaijmans e N. F. Lobo, «Gaps in protection: the actual challenge in malaria elimination», *Malar J*, vol. 22, fasc. 1, p. 46, feb. 2023, doi: 10.1186/s12936-023-04473-x.
- [37] «CDC - Malaria - Malaria Worldwide - How Can Malaria Cases and Deaths Be Reduced? - Drug resistance in the Malaria Endemic World». Consultato: 27 novembre 2023. [Online]. Disponibile su: https://www.cdc.gov/malaria/malaria_worldwide/reduction/drug_resistance.html
- [38] M. González-Sanz, P. Berzosa, e F. F. Norman, «Updates on Malaria Epidemiology and Prevention Strategies», *Curr Infect Dis Rep*, vol. 25, fasc. 7, pp. 131–139, lug. 2023, doi: 10.1007/s11908-023-00805-9.
- [39] P. Jagannathan e A. Kakuru, «Malaria in 2022: Increasing challenges, cautious optimism», *Nat Commun*, vol. 13, fasc. 1, p. 2678, mag. 2022, doi: 10.1038/s41467-022-30133-w.
- [40] D. Suandi *et al.*, «An evolutionary model propounding Anopheles double resistance against insecticides», *Applied Mathematical Modelling*, vol. 106, pp. 463–481, giu. 2022, doi: 10.1016/j.apm.2022.01.025.
- [41] A. M. Hammond e R. Galizi, «Gene drives to fight malaria: current state and future directions», *Pathogens and Global Health*, vol. 111, fasc. 8, pp. 412–423, nov. 2017, doi: 10.1080/20477724.2018.1438880.
- [42] G. K. Christophides, «Transgenic mosquitoes and malaria transmission», *Cell Microbiol*, vol. 7, fasc. 3, pp. 325–333, mar. 2005, doi: 10.1111/j.1462-5822.2005.00495.x.
- [43] T. Nolan, «Control of malaria-transmitting mosquitoes using gene drives», *Phil. Trans. R. Soc. B*, vol. 376, fasc. 1818, p. 20190803, feb. 2021, doi: 10.1098/rstb.2019.0803.
- [44] A. L. Smidler, D. G. Paton, G. M. Church, K. M. Esvelt, W. R. Shaw, e F. Catteruccia, «CRISPR-mediated germline mutagenesis for genetic sterilization of *Anopheles gambiae* males», *Synthetic Biology*, preprint, giu. 2023. doi: 10.1101/2023.06.13.544841.
- [45] R. D. Moreno, L. Valera, C. Borgoño, J. C. Castilla, e J. L. Riveros, «Gene drives, mosquitoes, and ecosystems: an interdisciplinary approach to emerging ethical concerns», *Front. Environ. Sci.*, vol. 11, p. 1254219, gen. 2024, doi: 10.3389/fenvs.2023.1254219.
- [46] B. L. Dickens, J. Yang, A. R. Cook, e L. R. Carrasco, «Time to Empower Release of Insects Carrying a Dominant Lethal and Wolbachia Against Zika», *Open Forum Infect Dis*, vol. 3, fasc. 2, p. ofw103, apr. 2016, doi: 10.1093/ofid/ofw103.

- [47] A. Hammond *et al.*, «Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance», *PLoS Genet*, vol. 17, fasc. 1, p. e1009321, gen. 2021, doi: 10.1371/journal.pgen.1009321.
- [48] E. Bier, «Gene drives gaining speed», *Nat Rev Genet*, vol. 23, fasc. 1, pp. 5–22, gen. 2022, doi: 10.1038/s41576-021-00386-0.
- [49] J. Li e J. Champer, «Harnessing Wolbachia cytoplasmic incompatibility alleles for confined gene drive: A modeling study», *PLoS Genet*, vol. 19, fasc. 1, p. e1010591, gen. 2023, doi: 10.1371/journal.pgen.1010591.
- [50] V. Bottino-Rojas e A. A. James, «Use of Insect Promoters in Genetic Engineering to Control Mosquito-Borne Diseases», *Biomolecules*, vol. 13, fasc. 1, p. 16, dic. 2022, doi: 10.3390/biom13010016.
- [51] M. Asad, D. Liu, J. Chen, e G. Yang, «Applications of gene drive systems for population suppression of insect pests», *Bull. Entomol. Res.*, vol. 112, fasc. 6, pp. 724–733, dic. 2022, doi: 10.1017/S0007485322000268.
- [52] Y. Kumam, H. N. Trick, P. V. Vara Prasad, e M. Jugulam, «Transformative Approaches for Sustainable Weed Management: The Power of Gene Drive and CRISPR-Cas9», *Genes*, vol. 14, fasc. 12, p. 2176, dic. 2023, doi: 10.3390/genes14122176.
- [53] K. Kyrou *et al.*, «A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes», *Nat Biotechnol*, vol. 36, fasc. 11, pp. 1062–1066, nov. 2018, doi: 10.1038/nbt.4245.
- [54] A. M. Hammond e R. Galizi, «Gene drives to fight malaria: current state and future directions», *Pathogens and Global Health*, vol. 111, fasc. 8, pp. 412–423, nov. 2017, doi: 10.1080/20477724.2018.1438880.
- [55] R. D. Moreno, L. Valera, C. Borgoño, J. C. Castilla, e J. L. Riveros, «Gene drives, mosquitoes, and ecosystems: an interdisciplinary approach to emerging ethical concerns», *Front. Environ. Sci.*, vol. 11, p. 1254219, gen. 2024, doi: 10.3389/fenvs.2023.1254219.
- [56] S. L. James, B. Dass, e H. Quemada, «Regulatory and policy considerations for the implementation of gene drive-modified mosquitoes to prevent malaria transmission», *Transgenic Res*, vol. 32, fasc. 1–2, pp. 17–32, apr. 2023, doi: 10.1007/s11248-023-00335-z.
- [57] E. Krzywinska, L. Ferretti, J. Li, J.-C. Li, C.-H. Chen, e J. Krzywinski, «femaleless Controls Sex Determination and Dosage Compensation Pathways in Females of *Anopheles* Mosquitoes», *Current Biology*, vol. 31, fasc. 5, pp. 1084-1091.e4, mar. 2021, doi: 10.1016/j.cub.2020.12.014.

- [58] E. Krzywinska, N. J. Dennison, G. J. Lycett, e J. Krzywinski, «A maleness gene in the malaria mosquito *Anopheles gambiae*», *Science*, vol. 353, fasc. 6294, pp. 67–69, lug. 2016, doi: 10.1126/science.aaf5605.
- [59] A. Simoni *et al.*, «A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae*», *Nat Biotechnol*, vol. 38, fasc. 9, pp. 1054–1060, set. 2020, doi: 10.1038/s41587-020-0508-1.
- [60] N. Page *et al.*, «Single-cell profiling of *Anopheles gambiae* spermatogenesis defines the onset of meiotic silencing and premeiotic overexpression of the X chromosome», *Commun Biol*, vol. 6, fasc. 1, p. 850, ago. 2023, doi: 10.1038/s42003-023-05224-z.
- [61] «Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data». Consultato: 15 febbraio 2024. [Online]. Disponibile su: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- [62] A. M. Bolger, M. Lohse, e B. Usadel, «Trimmomatic: a flexible trimmer for Illumina sequence data», *Bioinformatics*, vol. 30, fasc. 15, pp. 2114–2120, ago. 2014, doi: 10.1093/bioinformatics/btu170.
- [63] D. Kim, B. Langmead, e S. L. Salzberg, «HISAT: a fast spliced aligner with low memory requirements», *Nat Methods*, vol. 12, fasc. 4, pp. 357–360, apr. 2015, doi: 10.1038/nmeth.3317.
- [64] Consultato: 15 febbraio 2024. [Online]. Disponibile su: https://vectorbase.org/common/downloads/Current_Release/AgambiaePEST/fasta/data/
- [65] J. T. Robinson *et al.*, «Integrative Genomics Viewer», *Nat Biotechnol*, vol. 29, fasc. 1, pp. 24–26, gen. 2011, doi: 10.1038/nbt.1754.
- [66] H. Li *et al.*, «The Sequence Alignment/Map format and SAMtools», *Bioinformatics*, vol. 25, fasc. 16, pp. 2078–2079, ago. 2009, doi: 10.1093/bioinformatics/btp352.
- [67] P. Danecek *et al.*, «Twelve years of SAMtools and BCFtools», *Gigascience*, vol. 10, fasc. 2, p. giab008, feb. 2021, doi: 10.1093/gigascience/giab008.
- [68] Y. Liao, G. K. Smyth, e W. Shi, «featureCounts: an efficient general purpose program for assigning sequence reads to genomic features», *Bioinformatics*, vol. 30, fasc. 7, pp. 923–930, apr. 2014, doi: 10.1093/bioinformatics/btt656.
- [69] Y. Holtz, «Data visualization with R and ggplot2 | the R Graph Gallery». Consultato: 15 febbraio 2024. [Online]. Disponibile su: <https://r-graph-gallery.com/ggplot2-package.html>
- [70] M. D. Robinson, D. J. McCarthy, e G. K. Smyth, «edgeR: a Bioconductor package for differential expression analysis of digital gene expression data», *Bioinformatics*, vol. 26, fasc. 1, pp. 139–140, gen. 2010, doi: 10.1093/bioinformatics/btp616.

- [71] Y. Chen, D. McCarthy, P. Baldoni, M. Robinson, e G. Smyth, «edgeR: differential analysis of sequence read count data User's Guide».
- [72] C. Taxiarchi *et al.*, «High-resolution transcriptional profiling of *Anopheles gambiae* spermatogenesis reveals mechanisms of sex chromosome regulation», *Sci Rep*, vol. 9, fasc. 1, p. 14841, ott. 2019, doi: 10.1038/s41598-019-51181-1.
- [73] «The Gene Ontology (GO) database and informatics resource», *Nucleic Acids Res*, vol. 32, fasc. Database issue, pp. D258–D261, gen. 2004, doi: 10.1093/nar/gkh036.
- [74] P. D. Thomas *et al.*, «PANTHER: A Library of Protein Families and Subfamilies Indexed by Function», *Genome Res*, vol. 13, fasc. 9, pp. 2129–2141, set. 2003, doi: 10.1101/gr.772403.
- [75] M. D. Vibranovski, «Meiotic Sex Chromosome Inactivation in *Drosophila*», *J. Genomics*, vol. 2, pp. 104–117, 2014, doi: 10.7150/jgen.8178.
- [76] P. J. Wang, D. C. Page, e J. R. McCarrey, «Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse», *Human Molecular Genetics*, vol. 14, fasc. 19, pp. 2911–2918, ott. 2005, doi: 10.1093/hmg/ddi322.
- [77] A. Hammond *et al.*, «Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance», *PLoS Genet*, vol. 17, fasc. 1, p. e1009321, gen. 2021, doi: 10.1371/journal.pgen.1009321.