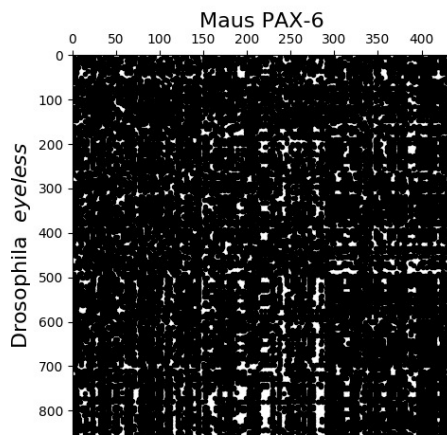


Aufgabe 1.1 Dotplot-Visualisierung

a)



- b) Die Übereinstimmungszahl (im Code als threshold oder thrsh bezeichnet) wird jeweils während des Durchlaufs der for-Schleifen über die x- und y-Achse pro Fenster ermittelt, mit der gegebenen Fenstergröße.

Eine dritte, innere for-Schleife (im fett gedruckten Abschnitt rechts enthalten) iteriert über die Indizes der Sequenzabschnitte in Fenstergröße (z.B. der String 'ABC' bei $n = 3$ -> drei Iterationen pro Fensterüberprüfung) und kalkuliert entsprechend des thresholds und der Fenstergröße, ob dieses Fenster gematcht wird.

CODE mit Übereinstimmungszahl / threshold:

```
def match(givenarray, thrsh):

    matcharray1 = []
    matcharray2 = []

    for indx, x in enumerate(givenarray[0]):
        for indy, y in enumerate(givenarray[1]):
            if x == y:
                matcharray1.append(indx)
                matcharray2.append(indy)
            else:
                thrsh_counter = 0
                for a, b in zip(str(x), str(y)):
                    if a == b:
                        thrsh_counter += 1
                if thrsh_counter / window >= thrsh:

                    matcharray1.append(indx)
                    matcharray2.append(indy)

    return [matcharray1, matcharray2]
```

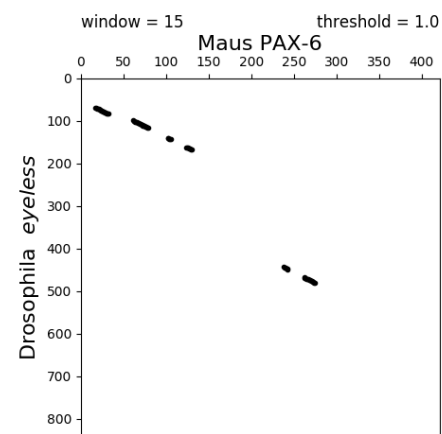
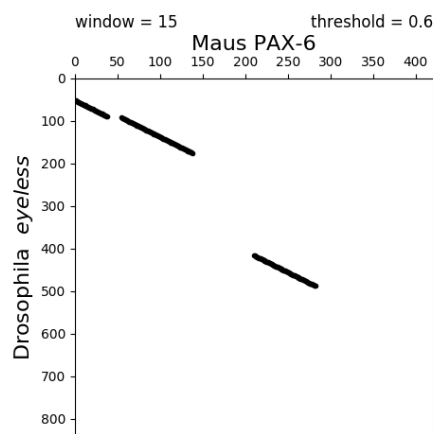
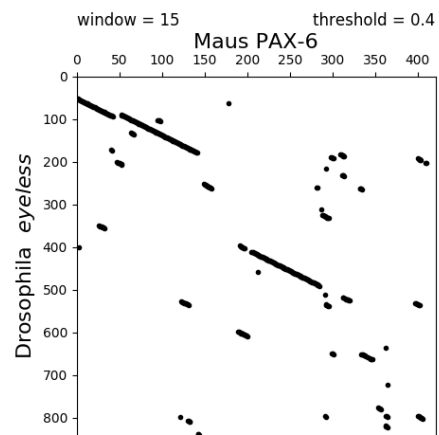
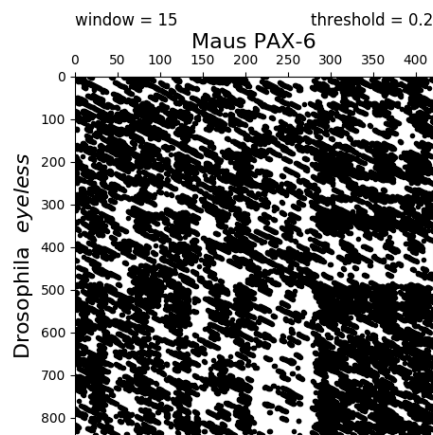
Bei einem threshold von z.B. 0,6 (also 60% Übereinstimmung) und einer Fenstergröße von 10 vergleicht die zip-Funktion die Strings der aktuellen Fenster Zeichen für Zeichen und inkrementiert pro Match einen Zähler (thrsh_counter), welcher dann durch die Fenstergröße geteilt und mit dem gegebenen threshold verglichen wird.

Daraus folgt: window: 10, threshold: 0.6, matches im Fenster:

7 -> $7/10 = 0,7$
 $0,7 \geq 0,6$ -> aktuelles Fenster wird gematcht.

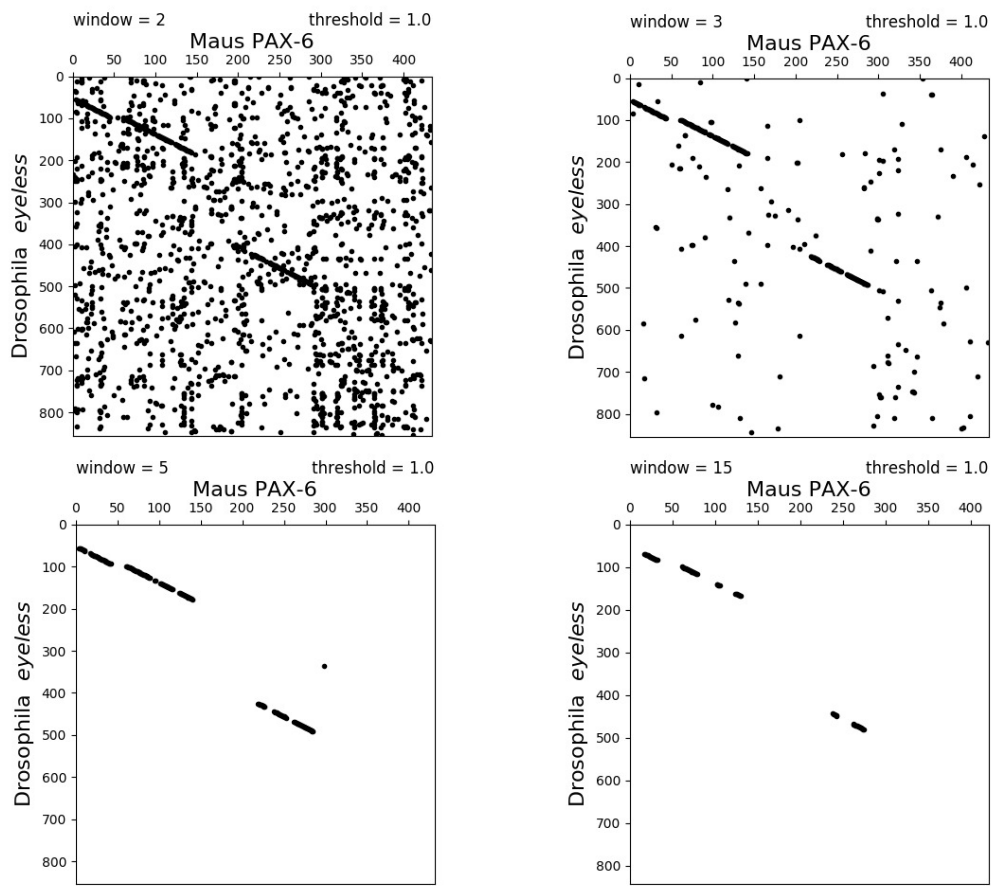
Zu 1.1 b)

Beispielbilder mit Fensterlänge 15 und verschiedenen thresholds (0.2, 0.4, 0.6, 1.0):



Kleinere Mutationen innerhalb einer Fenstergröße werden damit im Plot akzeptiert, wenn der Rest der Sequenz ansonsten hohe Übereinstimmungen aufweist. Somit dient der Threshold-Wert zur Sichtbarmachung von paarweisen Alignments trotz einzelner, mutationsbedingter mismatches.

c)



Die Ergebnisse der erstellten Dotplots stimmen optisch mit denen der gegebenen Plots mit verschiedenen Fenstergrößen ($n = \{2, 3, 5, 15\}$) überein.

Jeweils untereinander verglichen kann man feststellen, dass größere Fenster bei einer Übereinstimmungszahl von 1 (Match über gesamtes Fenster NUR, wenn beide Sequenzen komplett übereinstimmen) und mit einem Offset von 1 einen weniger verrauschten Plot ergibt und übereinstimmende Abschnitte deutlicher sichtbar sind.

Fragen, die gestellt worden sind:

Wie soll der Offset aussehen?

→ **Antwort per Mail:**

Selbst wählbar, empfehlenswert jedoch auf Wert 1.